

Fetuses and Parkinsonism

Concern about the use of fetal nervous tissue in the tentative treatment of Parkinsonism could easily be set to rest were not the underlying worry the use of abortion to remedy quite a different problem.

THERE are at least two views of the attempts made in Sweden, Mexico and (now) Britain to treat Parkinsonism by means of tissue grafts, either of adrenal tissue or nervous tissue. One is that the procedure may in due course be a means by which large numbers of people and their relatives are relieved of the symptoms of common and distressing conditions; Parkinsonism is not the only debilitating disease that might in due course be tractable. The other view is that the use of fetal nervous tissue provokes ethical doubts so serious that this route to therapy must be barred (as it has already been impeded in the United States — see page 668). What is the truth?

The procedures so far carried out to rid people of parkinsonism by transplanting tissue are not cures, but essays in that direction. The immediate cause of Parkinsonism is a lack of the neurotransmitter dopamine in the cells of the part of the brain known (on account of its colour) as the substantia nigra, but nobody knows whether this is the ultimate cause; most probably there are several. The fact that L-dopa, the precursor of dopamine, temporarily alleviates the symptoms is a sign both that Parkinsonism may be treated as if it were a deficiency disease and that there is more than that to be understood.

Attempts at therapy by tissue-graft are based on experiments with laboratory animals, and have shown relief of the symptoms of the animal analogue of Parkinsonism both by grafts of adrenal tissue (potentially a source of catecholamines, of which dopamine is one) and of fetal nervous tissue, distinguished from adult fetal tissue by the capacity of its cells to divide, possibly to differentiate according to the location in which they find themselves and perhaps more radically than L-dopa to remedy the underlying defect. But that is surmise only. Only time will tell what benefits will accrue to the patients. Only when there are large numbers of them will it be possible to judge whether either treatment has general value.

So why should there be a fuss in Britain that a surgical team in Birmingham has carried out two transplants of fetal nervous tissue? And why should Mr Robert Windom, assistant secretary for health in the US government, have instructed the National Institutes of Health not to continue in the field, using material from induced (as distinct from spontaneous) abortions, until an advisory committee on the subject has reported? There are (and certainly may be imagined) ethical problems in the use of fetal tissues for this or any other purpose. But there is also a close connection between these issues and the more contentious and political issue of induced abortion. Even in laconic Britain, there will be sparks flying when Mr David Alton's bill to reduce the time-limit for induced abortions comes up for debate in the House of Commons in a few weeks.

The ethical issues are easily listed. It would be wrong for a physician to procure an abortion to suit his patient's needs, but the remedy, which arises elsewhere in transplant surgery, is to erect a chinese wall between the physicians responsible for the patient and the donor, together with supervisory committees and even legal penalties for transgression. It would similarly be wrong for patients to conspire with their relatives to procure unwanted pregnancies, but the same chinese walls should suffice. Fears that, because transplant tissue must be as fresh as

possible, there will be a temptation to use "live" fetuses as sources are less substantial; fetuses should not be aborted so late that they are likely to be capable of independent life (which is the case for a modest reduction of the British 28-week time-limit on abortion). Why should the physicians acting for people with Parkinson's disease be supposed to differ from their professional colleagues in being monsters?

Windom's distinction between induced and spontaneous abortions is even more illogical. Under specified conditions, induced abortions are legal in most of the United States, as they are in many other countries. So what can be the moral difference between fetal tissue resulting from an induced and from a spontaneous abortion? Is the real question implicit in Windom's instruction, as in the British fuss, one of abortion? Having, twenty years ago, chosen abortion as the remedy for unhappy social problems, societies seem now to wish they could dispense with the remedy as well, especially now that many of the problems have gone away. It would be a pity if that subliminal connection were to dish a potentially valuable therapy. □

Greens against genes

... West Germany starts further back. Molecular biologists should stir themselves.

THERE may just be time to prevent the debate over genetic engineering in West Germany from turning into fisticuffs. Though the Bundestag first tackled the issue in 1984, the recommendations of the Enquete Kommission last year have set Bonn's ministries and committees the task of turning them into law (see page 672). With much of this going on in public, there is a danger that the debate will be a test of whose emotions are the more florid.

In the long run, everybody's interest is that the debate now joined about the proposed gene law should be conducted in a way that will not leave behind two exhausted armies of polemicists — one of them convinced it has been cheated of a deserved victory. But on present form there is just one army — that of those who say that genetic engineering is an abomination. Is it not time that molecular biologists (and their representatives and friends in organizations like the Deutsche Forschungsgemeinschaft) stirred themselves for long enough to demonstrate that there is another tale to tell? Temporarily, at least, they may have to resort to sloganeering: how about, for starters, "Is there a cure for AIDS without genetic engineering?"

West German researchers, self-effacing lest they seem otherwise, will not find this easy. Worse, they will have to win over politicians and electors at best unfamiliar with what genetic engineering is, but who have been persuaded by the past decade and a half of radical (and anti-intellectual) environmentalism that democratic prizes go to those who shout the loudest. The particular issue of genetic engineering, semantically linked as it is in West Germany with still recent horrors at places such as Auschwitz, is bound to be difficult. But that is merely another reason why the academic community should stir itself in its own interests. □



First ever animal patent issued in United States

- Harvard receives patent for *myc* mouse
- Furore from opponents of animal patent

Washington

THE US Patent and Trademark Office (PTO) last week granted the first patent in the world for a genetically altered animal. The patent was awarded to Harvard University based on the development in 1984 by Philip Leder at Harvard and Timothy Stewart, now at Genentech, of a mouse which readily expresses the *myc* oncogene, and is extremely susceptible to the development of tumours.

The *myc* oncogene occurs naturally in many animals, including man, and its genetic sequence is substantially the same in all species. Expression of the *myc* oncogene has been implicated in the development of breast cancer.

The Harvard patent's sweepingly broad main claim covers any transgenic non-human animal whose genome contains an activated oncogene sequence that has been introduced into the animal by recombinant DNA techniques. Additional claims specify that the animal could be a rodent, such as a mouse; that the oncogene sequence could be one that occurs naturally in the animal, but which is inserted with a different transcriptional promoter into a different place in the animal's genome; and that the oncogene sequence could be controlled by a promoter switched on by agents such as hormones. Animals which contain oncogene sequences controlled by the mouse mammary tumour virus promoter, the Rous sarcoma virus promoter and any synthetic promoter are specifically covered by the patent.

A patent covering a higher animal has been expected since last year when the PTO announced it would not refuse to issue a patent simply because the 'invention' was an animal. The PTO based its interpretation of patentable subject matter on the 1980 Supreme Court ruling in *Diamond versus Chakrabarty* that allowed the patenting of microbes, and several plant patent cases.

In the 1980 ruling, the Supreme Court decided that "everything under the Sun made by man" is eligible for a patent, as long as an invention is useful, novel and non-obvious, and the inventor describes the invention sufficiently to enable someone skilled in the art to practise it and enlarge upon it when the patent expires.

The PTO's announcement sparked a flurry of controversy over the ethical implications of designating inventors for higher life forms. For the past year, animal rights groups, farming organiza-

tions and religious leaders have spoken out against the new policy, and encouraged Congress to establish a moratorium on the patenting of animals until the ethical and economic consequences can be worked out. Farmers' organizations, backed up by anti-biotechnology activist Jeremy Rifkin, have been especially active in their lobbying efforts. A letter which leaked out of the PTO earlier this year stated that the reproduction of patented animals — such as improved breeds of livestock — would "seem to be an infringement" unless farmers paid royalties to the originators of the animals (see *Nature* 331, 651; 1988).

Last year, legislation to place a two-year moratorium on the patenting of animals was introduced into both houses of Congress, and nearly passed into law as an amendment to the 1987 budget supplemental appropriations bill. Legislation to enact a moratorium has been introduced into the House of Representatives again this year by Representative Charles Rose (Democrat, North Carolina). After the Harvard patent issued last week, Rose signed a letter along with 31 other representatives stating that "the patent office has been given no clear and certain signal from Congress... [that animal patenting] ...is acceptable public policy", and requesting that the PTO "immediately halt" the granting of additional animal patents until Congress has considered the issue.

Leder, Stewart and Paul Pattengale of the University of Southern California developed the '*myc* mouse' which led to the patent as a means to test whether the development of cancer results from the direct expression of one or more oncogenes, or whether additional agents such as mutagens or inducing hormones are required.

They constructed several genetic sequences containing the *myc* oncogene and promoter regions which would allow the gene to be expressed freely, and inserted them into fertilized mouse eggs. If only the *myc* oncogene was necessary for the development of cancer, cancer would have resulted uniformly in the breast tissue of the transgenic mice. Instead, tumours developed individually — but at a much higher rate than would be found in a normal mouse — suggesting outside substances are at work.

Harvard has granted an exclusive licence to Du Pont for the transgenic '*myc* mouse'.

Carol Ezzell

Killed HIV treatment in clinical trials

A NOVEL approach to curtail disease in patients already infected with human immunodeficiency virus (HIV), the virus that causes AIDS, is under way at the University of Southern California medical school.

The treatment was developed by Immune Response Corporation, formed last year by Jonas Salk, and involves injecting patients with HIV previously irradiated to destroy its DNA. Salk's theory is that injecting previously infected individuals with additional killed HIV will boost their immune response (see *Nature* 327, 473; 1987).

The trial is said to involve between 10 and 15 people who have early symptoms of AIDS. A similar post-exposure vaccine trial which involves the injection of killed simian immunodeficiency virus (SIV) in rhesus monkeys is being conducted by Paul Luciw at the University of California at Davis. Luciw reports that no immune side-effects have yet appeared in the monkeys, and that they have elevated antibody titres to SIV after six months and two booster injections.

The results of the Salk trial are expected to be presented at the Fourth International Conference on AIDS to be held in Stockholm, Sweden, this summer. C.E.

Ethical questions of tissue transplant

THE Department of Health and Human Services has put the brakes on plans of the National Institutes of Health (NIH) to transplant human neuronal tissue from induced abortions into patients with Parkinson's disease, a treatment already tried in Mexico (and in the United Kingdom, see p.667).

In a memorandum dated 22 March 1988 to NIH director James Wyngaarden, assistant secretary for health Robert Windom wrote that he was withholding approval of the proposed experiment and future experiments involving fetal tissue from induced abortions until NIH established an outside advisory committee to review such research.

Use of fetal tissue from induced abortions raises moral and ethical questions. Windom's memorandum raises several of these. By seeking informed consent from a pregnant woman, is a researcher in fact responsible for 'inducing' her to consent to an abortion? Windom wants the outside advisory committee to examine these and other issues.

This does not end research on fetal tissue at NIH. The memorandum does not include tissue from spontaneous abortions and stillbirths.

J.P.

Classified US technology for transfer to Japan

Tokyo

JAPAN and the United States last week signed an agreement that will allow transfer of information on classified US military technology to Japan. The agreement is one of several in a series designed to promote joint military research. But critics say that these moves, which mark a major shift in Japanese policy, are being made without public debate.

Under the new accord, signed by Foreign Minister Sosuke Uno and Mike Mansfield, the US Ambassador to Japan, Japan can gain access to classified US patents provided the information is used for defence purposes only and is kept strictly secret. Classified US inventions can be patented in Japan without public disclosure, and Japanese patents resulting from joint military research can also be classified. Furthermore, Japanese patent applications for technology that would expose classified patented technology will not be publicly disclosed unless it can be

established that the invention was made independently.

The agreement became effective immediately and does not require Diet approval because it invokes 'sleeping' provisions of a US-Japan agreement on interchange of patent rights and technical information for defence purposes which was approved by the Diet in 1956.

The agreement will smooth the way for Japan's participation in the US Strategic Defense Initiative (SDI), which Japan joined last year and will allow joint US-Japan development of other military technology.

Japan already produces military weapons under licence from US defence contractors but the technology is provided in the form of 'black boxes', the technology secrets of which are not disclosed to the manufacturers.

The new agreement is one of a series involving military research. Following the uproar last year over Toshiba Machine's illegal export of propeller-milling machines to the Soviet Union, Japan's Defence Agency and the US Pentagon agreed to increase cooperation in anti-submarine warfare research (see below).

And the new US-Japan Science and Technology Agreement, which is expected to be finalized in the next few months,

includes provisions to safeguard secret military research and prevent transfer of military technology or information to third countries.

In a recent television programme, Dr Koji Fushimi, former president of the Science Council of Japan and a member of the upper house of the Diet, expressed grave concern over Japan's increased commitment to joint military research with the United States. Japan spends only 0.01 per cent of gross national product on defence research, very much less than the United States, 0.7 per cent. Japanese researchers value their open research system and many economists argue that Japan's concentration on civilian research and development has helped the conquest of consumer markets.

But Fushimi fears that Japan's researchers will show weak resistance to the encroachment of secret military research. He says they should debate the issue more but there is a complete lack of information from the government on the new accords. Fushimi is particularly concerned about the new US-Japan Science and Technology Agreement.

When the original agreement was made in 1980 it was intended to promote joint research for 'peaceful' purposes. But the new agreement will be expanded to include military research, and problems over maintaining the secrecy of such research in Japan has been one of the major stumbling blocks in negotiations.

David Swinbanks

Who goes there?

Tokyo

THE narrow channels linking the Japan Sea to the Pacific Ocean are rumoured to be bristling with sophisticated listening equipment for detecting passing Soviet submarines. But one seldom hears about it — except, that is, when unsuspecting oceanographers encroach upon the area.

A joint team of researchers from Hokkaido University's Laboratory of Ocean Bottom Seismology and West Germany's Hamburg University are planning to carry out a seismic survey off the coast of Okinawa from Tuesday this week (19 April) by letting off powerful depth charges (up to 200 kg) at 130–150 m below the sea surface. The reflected seismic waves from the explosions will be picked up by ocean-bottom and land-based seismometers and will provide information on the structure of the Earth's crust and mantle down to 30–50 km in an area where the Philippine plate plunges beneath Japan.

But Japan's Defence Agency objected to the survey and asked the researchers drastically to modify their plans because the agency feared the 'artificial earthquakes' would affect their 'oceanographic equipment' installed on the seafloor.

After last-minute mediations by the Foreign Ministry, it has been agreed that the project can go ahead, provided the researchers give the agency warning of the timing and location of all explosions at least one day in advance.

David Swinbanks

UK relents on European space decision

London

THE British government announced this week that it will apply to participate in the Columbus programme of the European Space Agency (ESA), one of the 'Big Three' projects approved by ministers last November but which Britain had refused to support.

Britain now proposes to contribute £250 million to the £2,800-million, ten-year project in preference to taking part in the Canadian-led Radarsat venture, which would have cost Britain around £100 million and which will be four years in advance of Columbus.

Britain's interest in Columbus, which will form part of the international space station, lies principally in the polar platform, a remote-sensing satellite. When ministers agreed to the Columbus proposals (see *Nature* 330, 195; 1987) the projected cost of the polar platform was around £500 million. Pressure from Britain's space minister, Kenneth Clarke, resulted in a scaling-down of the specifications. The platform will not now be man-serviced and its weight will be reduced from 3.5 to 2 tonnes. The revised cost is some £315 million.

Britain is hoping that the bulk of the work on the polar platform will be awarded to British Aerospace. To allow Britain's involvement in Columbus, the government has allocated an extra £20 million to the British National Space Centre over the next three years, on top of its annual budget of £116 million. Clarke has also persuaded industry to provide £5 million to help offset the cost of participation.

A new director-general of the space centre was also announced last week. He is Mr Arthur Pryor, a civil servant from within the Department of Trade and Industry, the space centre's parent body. The centre has been without a director-general since the retirement in February of Mr Jack Leeming, who took over last September following the politically embarrassing resignation of Mr Roy Gibson in a gesture of exasperation at the government's refusal to meet an urgent request for more cash. Like Leeming, Pryor will be at Grade 3 on the civil service scale — a grade lower than Gibson when he occupied the director-general's chair. The change in the director-general's status is seen as an indication of the present administration's attitude towards space.

Simon Hadlington

US congressional committee takes on role of reviewer

Washington

A long and painful dispute over the accuracy of a paper published in the journal *Cell* became a public issue last week when a US congressional committee took evidence on the case. The dispute seems certain to gain notoriety as it raises fundamental issues of the rights of junior scientists to challenge their co-workers, public access to data on which published papers are based and the ability of scientific institutions to investigate charges of misconduct.

The paper, published in April 1986 (*Cell* 45, 247; 1986), was written by a team of six, including David Baltimore, a Nobel laureate working at the Massachusetts Institute of Technology (MIT). The other authors were David Weaver, Moema Reis, Christopher Albanese and Thereza Imanishi-Kari of MIT, and Frank Constantini of Columbia University. Imanishi-Kari was the principal author.

Questions about the paper's validity were raised in 1986 by Margot O'Toole, a postdoctoral researcher in Imanishi-Kari's laboratory. But little was heard in public until a private investigation of the case was attempted by Walter Stewart and Ned Feder, two National Institutes of Health (NIH) scientists known for their tenacious and controversial studies of scientific misconduct. According to their account last week, O'Toole came across 17 pages of records in Imanishi-Kari's laboratory which she recognized as being one part of the experimental data on which the *Cell* paper was based. After photocopying and analysing the data, she concluded that some of the principal conclusions of the *Cell* paper were incorrect.

O'Toole pointed out the discrepancies to the authors and urged that a correction be published. But this was not done, nor was O'Toole able to obtain the remainder of the data on which the paper was based. She took her doubts to Dr Henry Wortis of Tufts Medical School and later to Dr Herman Eisen of MIT. Both Wortis and Eisen produced reports on O'Toole's complaint. Neither was made public nor shown to O'Toole but both concluded there was no major error and that a published correction was unnecessary.

Soon after, O'Toole decided not to apply for further grants and abandoned her scientific career. Replying to questions at the congressional hearing, O'Toole said she had wished to continue in research, but as a result of her complaints she felt she had been portrayed as making "ridiculous, trivial and unsubstantiated complaints against my superior for vindictive reasons" and felt unable to do so.

The matter might have rested there if Feder and Stewart had not persuaded her

in 1986 to provide them with a copy of the laboratory records. From their own analysis, they concluded that the *Cell* paper "contained a number of serious misrepresentations of scientific fact". They stressed last week that they do not allege that there is any suspicion of fraud. But when they asked NIH for permission to publish an account of their analysis they found the weight of opinion against them.

Anonymous reviewers selected by NIH stressed that they had no way of assessing the validity of the laboratory notes obtained by O'Toole. As one put it, "the reader is placed in the same position as the receiver of an anonymous telephone call". Another pointed out that an uncontentioned explanation of the data is perfectly possible; for instance, that "the authors repeated their experiments elsewhere in their notebooks and found the experiment in question had been the subject of an interpretable experimental error".

NIH officials urged Feder and Stewart to contact the paper's authors directly. In response, Baltimore, who did not perform the research but has acted as the team's spokesman, pointed out that the paper had already been the subject of enquiries by "a number of respected immunologists". He rejected the idea that Feder and Stewart should set themselves up "as guardians of scientific purity" and conduct their own investigation of the paper.

Soon afterwards, Baltimore told NIH's deputy director of intramural research, Edward Rall, that if Feder and Stewart were still unsatisfied, the paper should be re-examined by independently chosen immunologists. The offer, described by Rall as "generous", was rejected by Feder and Stewart, who argued that Baltimore's wish that the decision of the committee be binding would restrict "free and open scientific debate".

After further pressure from Feder and Stewart, NIH decided they would after all permit publication of the analysis. But Feder and Stewart could not find a journal willing to accept it. Both *Cell* and *Science* turned it down, saying that an impartial committee of inquiry was the way to resolve the issue.

NIH's Mary Miers, head of the misconduct in science office, did indeed begin an investigation. But that too ran into trouble. Two of the first three scientists to be appointed had close links with Baltimore and later left. A new committee is now being formed and will, with luck, reach a conclusion later this year.

Committee members last week did not attempt to judge the scientific merits of the case but were clearly critical of the efficiency of procedures to deal with this

Criminal charge in scientific fraud case

Washington

IN what is thought to be the first federal criminal indictment for scientific fraud, Dr Stephen Breuning was charged last Friday with submitting false research results to support a grant application to the National Institute of Mental Health (NIMH). The offence was alleged to have taken place in 1983 when Breuning was at the University of Pittsburgh. Breuning denies the charges. If convicted, he could be fined and sent to prison.

Breuning's case was aired at the same congressional hearing investigating the possibility of error in Imanishi-Kari and colleagues' *Cell* paper.

Evidence was given by Robert L. Sprague, professor at the University of Illinois, whose allegations first brought the matter to the attention of the NIMH in 1983. He claimed that research reported by Breuning could not possibly have been performed in the limited time allowed for it. Breuning was researching on medicines given to mentally retarded people and the results of his experiments are said to have affected medical practice.

Despite the seriousness of Sprague's allegations, it took three years, and an article in the journal *Science*, for NIMH to produce a draft report on the case.

In the meantime, Sprague himself was investigated by NIMH. His research grant, which he had received from NIMH continuously for 16 years, was deferred in 1986. NIMH says there is no connection between Sprague's criticisms of NIMH practices and the deferment of his grant, but Sprague said in testimony to the congressional hearing that he seriously doubted that explanation. Alun Anderson

and other cases. There was particular concern that O'Toole felt herself to have been driven out of science and that NIH appears to do little to support those who allege error. Committee members were also critical of the very small resources NIH allocate to oversight. The committee chairman John Dingell expressed shock that NIH "relies completely upon the universities to investigate themselves".

The congressional committee did not, however, immediately suggest improvements. Some suggestions are that scientific fraud should become subject to criminal proceedings, and that scientists financed by public funds should be required to deposit their laboratory notebooks in public archives. But as Feder and Stewart themselves point out, the last thing that is wanted is "policing of scientific research by authorities outside of science". Alun Anderson

Australian scientists differ on how to attack fraud

Sydney

THE Australian scientific establishment has been thrown into confusion over how to investigate allegations that fraud was committed by a researcher at a Sydney-based research institute.

Accusations were made last December that Dr William McBride, director and founder of Foundation 41, a privately funded teratological institute, fabricated and altered data published in the *Australian Journal of Biological Sciences*. No one, it seems, is eager to investigate the case.

McBride is well known for his work on the link between the drug thalidomide and birth defects. In June 1982, he claimed in the published paper that orally administered hyoscine, an anti-cholinergic medication, causes birth defects in rabbits. An early draft of the paper appeared as evidence in a federal court trial in Orlando, Florida. The trial, in 1981, involved a lawsuit which alleged that Bendectin, also known as Debendox, contains an anti-cholinergic drug.

But the two junior scientists who conducted the experiments say that the data are false. According to Mr Phil Vardy, who has retained the original data, McBride added control groups and results, and altered the data to make it appear as though the defects were related to drug

dosage. The two scientists say that their names were added to the paper without their knowledge.

McBride claims that the extra rabbits were tested at the University of Virginia. But Norman Swan, producer of the Australian Broadcasting Corporation's science unit, whose programme "The Science Show" first made the allegations, has investigated further. He says that the University of Virginia scientist did not work with rabbits and had probably died at the time the testing occurred.

The matter did not cause comment until the *Australian Journal of Biological Sciences*, jointly published by the Australian Academy of Science and the Commonwealth Scientific and Industrial Research Organisation (CSIRO), received a second paper from McBride. The paper extended the initial findings and was preceded by a note correcting the original data table; it was published without being refereed.

Vardy and the other scientist involved, Jill French, wrote expressing their reservations and requesting a retraction of their names from the first paper. Their letter was never published. The editor-in-chief, Mr Paul Reekie, says that "the journal will remove the paper from its publication when and if fraud is proven".

It seems that the only avenue to investi-

gate the case is under the Medical Practitioners' Act which investigates medical, not scientific, improprieties. Foundation 41 says that it has invited an investigation by the Australian Academy of Science. The academy denies this saying it was asked to suggest a committee to look into the 'validity' of McBride's experiments. Jonathan Stone, a University of Sydney professor of anatomy and secretary of the academy, points out that "the issue is not whether the experiments are right or wrong but whether fraud has been committed".

The academy has said that even if invited to investigate the allegations, it would decline to do so, fearing litigation — both against those on the committee and those who present evidence. The National Health and Medical Research Council will not investigate the allegations as they do not support Foundation 41. Mr Terry Healy, a lawyer with CSIRO, claims that the organization has no legal right to investigate the matter.

The Australian Academy said last week that the only way to deal with cases of scientific fraud is through an independent body with the powers of a royal commission. No procedures at present exist to conduct such an investigation. **Tania Ewing**

Cyclosporin-yielding fungus found

New Delhi

THE chance discovery of a fungus at one of the Indian Council of Medical Research (ICMR) institutes has sent waves of excitement through the Indian medical and pharmaceutical community. While screening soil samples for mosquito-killing agents, scientists at the Vector Control Research Centre (VCRC) in Pondicherry have stumbled upon a strain of the fungus *Tolypocladium cylindrosporum* that yields large quantities of cyclosporin — an immunosuppressive drug widely used for managing patients undergoing organ transplants.

Cyclosporin is manufactured and marketed by Sandoz, which obtains it from the fungus *Tolypocladium inflatum*. According to VCRC microbiologist Dr K. Balaraman, a secondary metabolite in *T. cylindrosporum* produces an essentially similar product, as three-week-old fungal cultures. Animal studies are soon to begin to test the efficiency of the product which the VCRC is now able to produce in purified crystalline form.

The ICMR is now in the process of identifying an Indian company to undertake commercial production of cyclosporin. Several companies are keen to cash in on the VCRC director discovery which, according to Rajagopalan, "is perhaps the best thing to have happened in the Indian drug scene in recent years." **K. S. Jayaraman**

France moves towards laws on IVF

Paris

A working group of the French supreme court (Conseil d'Etat) has just presented its report to the government concerning medical ethics and the law. At the same time, Health Minister Michèle Barzach is publishing a set of decrees aimed at providing legislation controlling *in vitro* fertilization (IVF) and embryo research. Neither document is yet available from the government stationery office, but the French national daily newspaper *Le Monde* published extracts on 30 March.

At present, IVF and embryo research is limited only by guidelines published in December 1986 by a national consultative committee on ethics for health and the life sciences. There is no law specifically governing these activities. The ethics committee recommended proscription of IVF not specifically destined to help an infertile couple to have a baby. The committee also recommended the destruction of 'super-numerary' embryos only if the risk to the mother of allowing them to survive is great. Freezing the embryos was thought to be justified, to allow for the failure of embryos to implant in the uterus, but with a maximum shelf-life of 12 months. Regarding

embryo research, the committee recommended a moratorium of three years on genetic diagnoses using embryos before implantation. There is no state mechanism to ensure that these guidelines are followed.

The report from the Conseil d'Etat similarly sets out the areas where legislation is necessary, in a broad spectrum of medical practices and experimentation on human subjects. The committee would like any clinical trials that do not have therapeutic aims banned and to ensure that the full and active consent of the participants is given. These concerns are given added weight by the case of Dr Milhaud who performed an anaesthetics experiment on a 'brain-dead' patient without any consent or therapeutic goal.

Barzach aims to limit the abuse of genetic engineering techniques and clinical trials on humans when there are no therapeutic aims. But the decrees have been criticized as a stopgap measure when full legislation is required — and desired by doctors and biologists. The Conseil d'Etat also recommends the drafting of legislation, but Barzach, says *Le Monde*, feels that this will be a matter of years. **Peter Coles**

Molecular biology threatened by West German law

Munich

AFTER years of internal debate, the West German government is considering legislation that may jeopardize both research and production using recombinant DNA technology. Two laws now being drafted seem certain to have far-reaching, if not devastating, effects.

West Germany's broad-based environmental movement, including but not restricted to the Green Party, is preparing for a fight against genetic engineering that promises to be as intense as the continuing struggle against nuclear energy. There has already been a bomb threat against the Heidelberg Molecular Biology Centre (ZMBH).

Opinion has been crystallized by the report of the 'Enquete Kommission' appointed in 1984 by the Bundestag (parliament) to study the "Prospects and Risks of Genetic Engineering". In January 1987 the report urged more than 170 specific measures (see *Nature* 325, 474; 1987) covering such areas as cloning of human beings, release of genetically manipulated microorganisms and genome analysis by employers and law enforcement agencies. The Bundestag has begun hearings on some of these issues, which are meant to lead to a new 'gene law' by early next year.

Meanwhile, another measure on waste materials produced by genetic engineering has provoked a sharp response from the few researchers who have heard about it. (The Deutsche Forschungsgemeinschaft, which supports most university basic research, will probably file an objection soon.) It consists of a set of regulations to fill a gap in the West German Water Management Act (*Wasserhaushaltsgesetz*).

The first draft of the regulations would handle any plasmids or vectors containing foreign DNA as if they were pathogens and require that the bacteria be chemically or thermally treated "to destroy all materials capable of reproducing" outside the laboratory or production facility.

The measure was provoked by the pharmaceutical company Hoechst AG's plan to produce human insulin using genetically manipulated *Escherichia coli* bacteria. Although the Hoechst plant near Frankfurt was initially approved by local authorities (see *Nature* 329, 193; 1987), plans have now been stopped by a citizens' initiative, the 'Höchster Schnüffler und Maagucker'.

Molecular biologist Ernst-Ludwig Winnacker, director of the Munich Gene Centre, argues that the wastes should be treated according to their pathogenicity. Hoechst says that the insulin-producing bacteria are indeed incapable of surviving

in the outside world, but the principal author of the proposal, Jürgen Hahn of the Federal Health Office (Bundesgesundheitsamt), says that the law is meant to anticipate "not the pathogenic but rather the ecological consequences" of an accidental release. The operative principle is that, once released into the environment, new genes cannot be brought back.

The same arguments can be expected as the Bundestag nears passage of a gene law. Moderates such as Bundestag member Wolf-Michael Catenhusen (Social Democrat), who chaired the Enquete Kommission, consider it likely that the law will give legally binding status to the recommendations of the Zentrale Kommission für Biologische Sicherheit (Central Commission for Biological Safety), which can now propose guidelines only for federally supported research.

Researchers would support such a scheme as long as the regulations governing genetic engineering are not specified in the law. Otherwise, says geneticist Peter Starlinger of the University of Köln, changes in the regulations would require Bundestag approval, bringing delay and preventing West Germans from keeping up with advances in the field.

Environmentalists would feel betrayed by such a plan. In the view of Frankfurt attorney Hans Neumann, the ZKBS is inherently biased because some members come from private industry. (The ZKBS has twelve members, including six biologists and three representatives of industry, among them one man from Hoechst.) Furthermore, "the professors are not free in their decisions". The tone of the debate may have already been set by the opponents of genetic engineering, who have begun to use the slang word for a nuclear accident ("GAU") for an accidental release of recombinant DNA.

The bomb threat at the ZMBH may be a foretaste of what is to follow. The building, then under construction, was damaged by a bomb in 1985, but none was found this time.

Researchers cite one-sided media coverage as a cause for the public fear of genetic engineering. "Even scientists and physicians are being made afraid", says Hermann Bujard, director of the ZMBH. ZKBS member Gerd Hobom of the University of Giessen adds that "in the United States, there is only one Jeremy Rifkin, but here there are lots of him and they are better organized". If the laws materialize in their most extreme form, Bujard says that "we'll have an exodus" of molecular biologists: "all the good students will go to the United States." **Steven Dickman**

Frog liberationists croak up

Berkeley

ANIMAL rights has moved into the classroom in California, where a new law gives primary and secondary school students the right to refuse to perform animal dissection, provided there are alternative means of studying the material.

The law, the first of its kind in the United States, is likely to encourage more students to protest against classroom dissection. The issue came to national attention when Apple Computer ran television advertisements featuring a California high-school student who refused to dissect a frog, proposing instead to learn from a computer program called 'Project Frog'. The California law was drafted at the request of the Peninsula Humane Society in San Mateo, one of the leading animal rights organizations in California. Critics say the law creates an issue where there really is none: few California students have opposed dissection, and, according to a California School Board Association spokesperson, the cases have been handled adequately by local school officials. The new law will not actually force schools to excuse students from dissection, rather it will leave the ultimate decision up to the individual school. As such, it is not a state mandate, said one school board official, and does nothing to change the way schools now handle such cases.

But it promises to increase the number of cases the schools must address. Supporters of the law say its main purpose is to draw more attention to the plight of animals that are dying needlessly in the name of education, and to encourage students to demand alternatives.

At least one other state may soon follow California's lead. The Massachusetts legislature is currently considering a similar bill. **Marcia Barinaga**

Maths prize

London

THIS year's Crafoord prize has been won by Brussels-born Professor Pierre Deligne of the Institute of Advanced Study in Princeton, New Jersey, and Berlin-born Professor Alexandre Grothendieck of the Université des Sciences et Techniques du Languedoc in Montpellier, France. It was the turn of mathematics for the prize, which is awarded by the Royal Swedish Academy of Sciences in areas not covered by the Nobel prizes. The two mathematicians are cited for their fundamental research in algebraic geometry, and will share an award of \$270,000. Swedish research in the same area will benefit from a subsidiary award of \$150,000. **Peter Newmark**

Government departments clash over US drilling proposals

Berkeley

THE Environmental Protection Agency (EPA) has joined a growing chorus opposing the Interior Department's plan to lease 1.1 million acres off northern California for oil and gas exploration.

Interior Secretary Donald Hodel proposes leasing more than 200 three-mile-square tracts as the first step of a five-year plan to develop northern Californian oil and gas. Optimistic estimates predict as much as 790 million barrels of oil. But

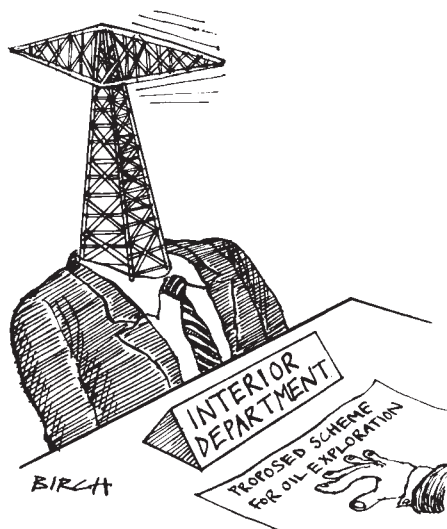
opponents of the plan like to point out that even such optimistic estimates would satisfy US demand for fuel for less than two months.

Supporters of the oil exploration, including the California Chamber of Commerce, argue that the state has a responsibility to allow development of resources off its coast.

The Interior Department issued a draft environmental impact statement last December, and is currently reviewing public comments on the statement. If it goes through with leasing, the tracts will be auctioned in February 1989.

In its response to the impact statement, EPA expressed concern for two national wildlife refuges, the Farallon Islands, about 30 miles off the Golden Gate, and Point Reyes, on the coast north of San Francisco. The refuges are important breeding grounds for 10 species of seabirds, including the common murre, a diving bird particularly sensitive to oil spills. The area is also of importance to fisheries, and supports many marine mammals, including migrating whales, says EPA spokeswoman Virginia Donohue.

EPA has asked the Interior Department to find leasing configurations more protective of the Farallons, Point Reyes and several river-mouth habitats.



Spain lays down the law on research plans

Barcelona

THE National Plan for Research (NPR) has finally been approved by the Spanish government. It represents the main attempt by the Socialist government to put into effect the Law of Science of 1985 and to concentrate the effort devoted to civil research. The plan covers the period 1988–91 and is intended to co-ordinate the programme of the various ministries involved in research, and to increase expenditure on research from 0.7 per cent of gross national product to 1.2 per cent by the end of 1991. It is also proposed to increase by around 50 per cent the number of researchers, who now represent 0.12 per cent of the population.

The NPR requires the expenditure of more than 600,000 million pesetas (£3,000 million) in four years — though only a third of this represents actual funding of research in Spanish government and industrial laboratories. The budget also has to cover Spanish participation in international programmes and administrative expenses. The plan includes programmes aimed at encouraging high-quality basic research and at stimulating international co-operation.

But the most important element is a group of 23 programmes where efforts will be concentrated in defined areas. Production technology (microelectronics, robotics, new materials and so on) and telecommunications account for half the total spending. The 'natural resources and agrobiology' programmes include food technology, agriculture and geological resources; it also includes a programme of Antarctic studies as a continuation of a recent expedition on which for the first time a Spanish base was occupied during the Antarctic summer. Another important programme is biotechnology, in which most of the funding will go to the new Centre for Biotechnology in Madrid. In the opinion of the secretary general of the NPR, Dr Emilio Muñoz, the plan provides insufficient funds for social and human sciences and for health research, but annual revisions of the plan will make it possible better to define and finance the weakest areas. Another imponderable is the response of private industry. Few Spanish companies have so far shown an interest in adopting new technologies, and unless things change, both the results of the research and the trained personnel may be wasted.

Pedro Puigdomènech

The EPA recommendations are not binding, but as a federal agency, EPA expects to be taken seriously. Its comment is one of 5,000 the Interior Department has received from federal and state agencies, environmental groups, industry and thousands of individuals. Drilling opponents have also packed the house at several volatile hearings on the plan. Although it seems likely that oil development will proceed, the department may be forced to modify its plan in response to the public comments.

Marcia Barinaga

Universal health insurance in Massachusetts

Boston

A landmark bill passed into law last week makes Massachusetts the first state in the United States to provide a universal health insurance plan. Under the scheme, employers, workers and state taxpayers will gradually assume financial responsibility for health care for uninsured Massachusetts residents.

The bill, which received narrow approval in the state senate after a marathon 12-hour debate and a tumultuous nine-month legislative history, seeks to provide health care coverage for the 600,000 of the state's 5.8 million residents who are currently uninsured. Half the potential beneficiaries are also unemployed.

Under the act, companies with more than five employees must pay a yearly surcharge of roughly \$1,680 per worker which will go to a new health insurance fund. If, however, these companies offer health insurance benefits to their workers, they will be allowed to deduct the cost of insurance from the surcharge. Small businesses with fewer than five employees are exempt. Starting in 1992, the new law will allow uninsured residents to pay for their own insurance on a sliding scale based on how much they earn.

Michael S. Dukakis, Massachusetts governor and the current leader in the race to be the Democratic candidate for president, was a strong supporter of the bill and made it a prominent topic in his presidential bid. Massachusetts Republicans complained that the bill had been pushed through the state legislature especially quickly in order to influence the presidential campaign.

But the measure has generated widespread support in Massachusetts, and other state legislatures will be watching its implementation closely. The bill is expected to cost the state approximately \$1,000 million over the next four years, and to generate roughly the same amount in additional revenues to the local hospital industry.

Seth Shulman

Anthrax outbreak in Soviet Union due to natural causes?

Washington

AN outbreak of anthrax in the Soviet city of Sverdlovsk nine years ago is still a hot topic for debate. Last week, three Soviet health officials presented evidence at a forum at the US National Academy of Sciences that the anthrax outbreak was caused exclusively by contaminated meat. But the United States continues to insist that the true cause of the epidemic was an explosion at a biological weapons facility near Sverdlovsk that released anthrax spores into the air.

The United States has repeatedly pointed to the Sverdlovsk incident as an example of Soviet non-compliance with biological weapons agreements. The Pentagon view is that after spores were released, the Soviet Union attempted to neutralize them by aerial spraying with a hypochlorite solution. There were also a large number of military casualties and heavy military involvement in controlling the incident, according to the Pentagon.

But the Soviet health officials presented a very different picture. Before an audience of epidemiologists and other US health researchers gathered at the academy, they attempted to demonstrate that the hypothesis that the Sverdlovsk incident resulted from airborne spores was inconsistent with the facts.

Pyotr Burgasov, former Soviet deputy minister of health, said there were 96 cases of anthrax related to the Sverdlovsk incident. Of these, 17 exhibited the generally non-lethal cutaneous variety, and 79 the more dangerous but rarer gastrointestinal form, of whom 64 died.

Burgasov was among the first health officials on the scene after the outbreak occurred. He said there were no military involved in controlling the epidemic, no aerial spraying for decontamination or any other purpose and no evidence for the type of anthrax that would result from the inhalation of airborne spores.

Instead, Burgasov argued the incident

could be explained by consumption of contaminated meat. A combination of high demand for meat and dwindling supplies of animal fodder drove area farmers to make use of bone meal. Anthrax is endemic to the Sverdlovsk area, and Burgasov explained that some of the bone meal may have been contaminated, thus infecting meat supplies. He said Sverdlovsk residents were urged to turn in any meat they had purchased and were storing, and in many cases anthrax spores were isolated from the meat. Burgasov admits that there is a military facility at Sverdlovsk, but said it was not involved in the incident.

The Soviet visit was organized by Harvard University biologist Matthew Meselson, who acted as an adviser to the United States government in 1980 when it first sought to determine whether the Sverdlovsk incident could be the result of

a biological weapons effort gone awry. At the time, Meselson felt the evidence was equivocal, and he spent the subsequent years trying to find the appropriate Soviet officials who could provide a more complete explanation.

Philip Brachman, an epidemiologist at Emory University and former chief of epidemiology for the Centers for Disease Control (CDC) in Atlanta, agrees that it was difficult to form a conclusion based on evidence available in 1980. After hearing the explanations provided by Burgasov and his colleagues, Brachman agreed that the anthrax outbreak did not seem to result from an airborne disease, although he pointed out that the Soviets had not presented all their raw data. Alexander Langmuir, also formerly a chief epidemiologist at CDC, went further, saying that the US position needed to be reviewed in the light of the Soviet evidence.

But by the end of last week, the Pentagon had not budged. According to a spokesperson, the Soviet Union "has yet to provide a satisfactory explanation of the event".

Joseph Palca

Therapeutic proteins overtake monoclonals

Slough

HALF a kilogram of protein would not feed a multitude but the profits of selling that quantity of some proteins would. While Celltech was feeding a moderate multitude at the opening of its modish new laboratory and office building last week, half a kilogram of a protein of considerable value was being brewed next door in the first production run of Celltech's 2,000-litre fermenter.

Monoclonal antibodies, of which Celltech is a master brewer, can sell for \$1 million per kg and the company can now produce 15 kg per year. But the demand for monoclonals, at least for diagnostic purposes, is less than had been anticipated.

As a consequence they are likely soon to become only a minor component of the

proteins made in Celltech's fermenters. As a sign of the times, the new fermenter's debut was assigned to production of a cell-surface protein, whose further identity was not forthcoming, rather than a monoclonal.

Another sign was the pull-out earlier this year of the Boots Company from Boots-Celltech Diagnostics Ltd, after only two years. One reason for this was last year's loss of £2.45 million of the joint venture. Nevertheless, claims Celltech's chief operating officer, David Gratton, there is no shortage of suitors to take over the stake held by Boots.

The demand for therapeutic proteins is, however, growing. Celltech has already drawn up provisional plans to install a 10,000 litre fermenter. Its construction would be demand-led, say company personnel, hinting that the demand may soon arise from a contract that is under negotiation. Beyond that, sights might be set on a separate production plant, perhaps in a different locality.

Celltech's need to produce larger quantities of protein grows at a rate that depends partly on how many contracts it wins, and partly on the speed with which its own products find a market. Confirming that it has decided to create its own business in biopharmaceuticals, including marketing and sales, Celltech's recently published 1987 Annual Report suggests two early products. One is a hybrid of an enzyme that dissolves blood clots and an antibody that delivers it to the clots; the other is 'calcitonin gene-related peptide', which may be used in the treatment of cerebrovascular problems.

Peter Newmark

A further LINK

London

BIOTRANSFORMATION is to be the sixth area of collaborative research in the UK Department of Trade and Industry's LINK programme (see *Nature* 331, 473; 1988). At least 15 companies, including many household names, are expected to contribute a total of £2 million over three years, with matching funds from the department together with the Science and Engineering Research Council. At least seven universities will be involved, with a focus on Kent, Essex and Warwick.

Peter Newmark



Raising the roof — an existing building extended to accommodate Celltech's new research centre.

Nothing held back in Indian – Soviet scientific collaboration

New Delhi

THE long-term Indo-Soviet science agreement signed last July got off to a start last week with the two countries agreeing to work on 70 specified projects in eight broad areas. In the next two years, the Academy of Sciences of the USSR will throw open 90 of its institutes to 200 visiting Indian scientists who will stay up to 10 months. India will reciprocate. According to the academy president, Guri I. Marchuk, "the USSR has no science programme with any other country in the world on such a large scale".

By involving top scientists in the programme, the Soviet Union is trying to show that it is viewing the collaboration very seriously. "When we say our labs are open to Indians, we will not hold back any secrets", said an academician. To prove this point, the Soviets have agreed to share with India two of its advanced technologies: the manufacture of diamond films for coating materials and the induc-

tion of thermonuclear fusion.

Among the identified projects, half belong to the area of basic science and one-third to product-orientated technologies such as catalysts, superhard materials and special lasers for surgery.

The rest are concerned with futuristic technologies from which both countries will benefit commercially. The cooperation will work at three levels: exchange of scientists, creation of scientific centres jointly manned by experts from both countries, and sharing of information through seminars and symposia.

The Soviet Union will establish a facility in India for the production of oral polio vaccines. The Soviets will also set up a power metallurgy centre in Hyderabad that will broaden research in new materials, and the two countries will exchange information and samples in the area of high-temperature superconductors.

The Soviet Union will help the Bhabha Atomic Research Centre in the design, construction and operation of the proposed synchrotron radiation source at the Centre of Advanced Technology in Indore which has plans to start building a 20-MeV linear accelerator by 1990. The Soviets will also help to set up facilities for making laser glasses and growing shaped crystals, and will assist in making gas lasers for medical applications. India also proposes to buy a continuous-wave carbon-dioxide laser for material processing.

Although India will benefit most from this collaboration, the Soviet scientists are dismayed by the lukewarm response from their Indian counterparts. The reasons for this, presumably, are the language barrier and unfamiliarity with a cold climate.

Another disincentive, according to an Indian participant, is money. Indians sent to the Soviet Union under this agreement will get just 18 roubles a day for short stays or 450 roubles per month for extended stays to support their families. Their living allowance for travel to the West is four times higher.

K.S. Jayaraman

Magnetic appeal

London

A NEW nuclear fusion experiment, designed to test superconducting technology for fusion reactors, has started up in France. The TORE-SUPRA reactor, at the Cadarache Centre near Aix-en-Provence, uses the most powerful superconducting magnets yet tried on a fusion tokamak apparatus.

The large doughnut-shaped reaction vessels of tokamaks conventionally use copper-wound magnets to confine the hot plasmas (about 10^8 K) in which light nuclei fuse and release energy. It is thought, however, that commercial reactors will need superconducting magnets to sustain the high fields for useful durations. TORE-SUPRA, a joint venture of Euratom and the Commissariat à l'Energie Atomique (the French atomic energy commission) costing FF800 million, is a testing bed for the necessary techniques. When it reaches full specification in the next two years, TORE-SUPRA will generate a magnetic field of 45,000 gauss, maintaining a plasma carrying a current of 1.7 million amperes for 30 seconds. These will be the first studies of tokamak operation in the steady or quasi-steady regime.

A similar Soviet experiment, T-15, is still under construction. These and current conventional experiments in Europe (JET), the United States (TFTR) and Japan (JT-15) will provide the basis for the international construction of a prototype commercial fusion reactor, ITER, in the 1990s.

Roland Pease

EEC decision on fusion delayed as Britain cries "too costly"

London & Brussels

THE European Community's proposed £4,000 million, five-year research programme inched further towards realization last week with the provisional or final endorsement of another batch of projects by ministers meeting in Luxembourg. But Britain once again incurred the anger of its fellow members by complaining at the proposed 911 million ECU (European currency unit; 1 ECU=£0.69) budget for the nuclear fusion programme. There is also disagreement about how to reorganize the Community's Joint Research Centre, which is scheduled to receive 700 million ECU until 1991. The ministers are due to meet again in late June to resolve the outstanding issues.

Agreement was reached on some 2,000 million ECU of projects. The only project to win final approval was the 1,600 million ECU second phase of the ESPRIT programme of information technology, now being promoted as the flagship of the Community's joint research endeavour. ESPRIT II will concentrate on micro-electronics, information processing systems and information technology applications, with 'new emphasis' being placed on computer-integrated manufacturing (CIM) and application-specific integrated circuits (ASICs).

Common positions — Community jargon for provisional agreements — were achieved on several other projects. The Science programme to promote interaction within the Community is to receive

167 million ECU. DRIVE and DELTA, offshoots of ESPRIT and its sister telecommunications programme RACE, are concerned with information technology applications for road traffic and distance learning; they will receive 60 million ECU and 20 million ECU, respectively. A common position was also adopted on the revision of the 1985–89 biotechnology programme, supplementing its budget of 55 million ECU with a further 20 million.

The difficulty over the fusion programme has resulted from disagreement about how much of the proposed budget should be held over beyond 1991. This principle of 'overhang' was instituted as the price for Britain's agreement to the overall Framework research programme, agreed in essence last October, after a nine-month delay. The European Commission has proposed a fusion budget of 911 million ECU. For Britain, the overhang is too small — it wants it to be 100 million ECU and has indicated that it is not prepared to shift. At last week's ministerial meeting, a compromise figure of 35 million ECU was put forward, which received the support of 11 of the 12 delegates.

Most of the Community's fusion work is done at the Joint European Torus, at Culham in England, where there is growing despondency at ministers' failure to agree a budget. As things stand, the laboratory is having to work on a reduced budget of a monthly allocation of one-twelfth of last year's spending.

Simon Hadlington & Bernard Conlon

The DNA data libraries

SIR—A leading article in your journal recently commented on the sequencing of the human genome (*Nature* 331, 465; 1988) after a report from the US National Research Council on this topic. First, *Nature* recommends that this project should be an informal part of the scientific endeavour without any coordination and in the next column the author paradoxically suggests that an internationally coordinated mega-databank should be introduced. More careful research would have revealed that the National Institutes of Health (NIH), the European Molecular Biology Laboratory (EMBL) and the Japanese funding agencies have taken several steps towards this goal. They have also recognized the inevitable obsolescence of the current databases and, with this in mind, began research into novel approaches with an international workshop as long ago as February 1987.

A unified nucleic acid sequence database may be available within 2–3 years, based mainly on data transmitted directly by researchers and integrated with other data collections such as protein data, physical and genetic gene maps and protein 3-D structures. This collaboration involves Europe, the United States and Japan and 50 per cent of the submissions already come from the originators of the sequence information, that is, the scientists themselves. Three days after your leading article was published, the international advisory board of the databanks, comprising three experts from both the United States and Europe and two from Japan, met in Bethesda, Maryland, to discuss priorities for future tasks and to review collaborative projects between the databanks. The recent collaboration in a number of areas, including the adoption of a common data entry form and the successful launch of a direct submission system, was recognized and praised. The advisory panel finally strongly recommended that the outstanding problem in unifying some of the more crucial aspects of annotation should be solved in the next 6–9 months. Integration with other kinds of databases should also be pursued as rapidly as possible.

The procedure introduced on 1 January 1988 in a collaboration between the journal *Nucleic Acids Research* and the EMBL data library requests authors to submit sequences to the EMBL data library in computer-readable form, before journal submission and review. This will not only speed up the entry of sequences into the banks but will also provide an entirely new type of publication. It is desirable that this approach be accepted by all journals in the field. The advisory board therefore expressed a strong recommendation that the scientists should be

primarily responsible for the data presented, with any information added by the databank or other experts being clearly separated from the original submission.

Additional resources have recently been made available to the data libraries. Genbank has received a \$17 million contract from NIH for the next 5 years and EMBL has, through the establishment of a science-led biocomputing programme, provided the hardware and software base. EMBL has also requested funds from the European Community and the European pharmaceutical and chemical industry for establishing a European network with regional nodes to coordinate the anticipated traffic of nucleic acid sequences.

It suffices to conclude that the anomaly with the databanks is on its way to being rectified and that we have already come a long way towards taking care of the data from a human genome project if launched. We are close to a joint international data source in biology. International coordination in science is more dependent on hard undisturbed work than on bombastic and uninformed statements. *Nature's* journalists would profit from doing their homework before formulating their regretfully influential and sweeping statements.

LENNART PHILIPSON
European Molecular Biology Laboratory,
Postfach 10.2209,
Meyerhofstrasse 1,
6900 Heidelberg, FRG

Phd by thesis

SIR—Beverly Halstead's swingeing attack (*Nature* 331, 497; 1988) on established practice may focus attention on an important issue, but he goes too far in equating the doctoral degree with publications. Is he seeking to fit untidy researchers into a neat package of academic measurability?

A firm distinction must be made between (1) aptitude for research (the PhD degree) and (2) an end-product of research (publications). There are numerous reasons why the results of much painstaking research fail to appear in print. Negative results are notoriously difficult to publish and poor use of English may be a bar to publication while casting no light on research aptitude. (And publishing houses are more amenable to economic and political pressures than are degree-awarding bodies.)

But approaches of novelty are often suppressed purely for their originality. Although most students may fit into Halstead's box, what of those individuals whose radical thinking becomes an inspiration to us all? It can be quite acceptable to award a PhD degree even if one suspects that the conclusion of the disserta-

tion is wrong: scholarly disagreement alone does not constitute grounds for refusal. The opposite is true for publication.

In fairness to Halstead's argument, if good publications are indeed an adequate expression of research aptitude, a dissertation might well consist of a group of publications bound together with a general introduction and conclusion (raising, of course, the question of what precisely constitutes a publication). Unfortunately, it is often all too easy for the mediocre student in a dynamic research group to accumulate a series of publications.

A publication record can be neither a necessary nor a sufficient condition for the award of the PhD degree. May it not be that the tried and tested method of thesis submission and review, in good faith, by a panel of established researchers is so far the only adequate procedure for assessing aptitude for research?

RICHARD LATHE
CNRS-LGME & U184-INSERM,
11 Rue Humann,
67085 Strasbourg Cedex,
France

SIR—The PhD degree has three main variables, the student, the supervisor and the project. It is the supervisor who can influence the current trend in "ever-expanding monsters". The writing of a PhD thesis should be an enjoyable experience if the student receives guidance on its proper function and preparation.

T.K. BRADSHAW
Sittingbourne Research Centre,
Sittingbourne, Kent ME9 8AG, UK

US budget deficit

SIR—Your advice ("Transatlantic tax dilemmas", *Nature* 331, 643; 1988) to increase US income tax implicitly assumes that the US Congress will use the expected revenue increase to reduce the deficit. Nothing could be further from the truth. There are numerous proposals now before that august body to appropriate all kinds of goodies for special interest groups, ranging from free day-care centres down to funding of airports for billionaire playboys. Any additional tax receipts will go toward those expenditures, not to reducing the deficit. A year or so ago, a study was conducted to determine the experience from past 'solutions' of the proposed type. It was found that the deficits did not go down; they went up. The truth is that the current US Congress is a deeply flawed institution run by professional politicians interested in only one thing: reelection. In the words of the immortal Vince Lombardi, "Winning isn't everything; it is the only thing!"

ROBERT C. WHITTEN
1117 Yorkshire Drive,
Cupertino, California 95014, USA

One-dimensional ice made clear?

A dynamical solution of the behaviour of a crude one-dimensional model of ice should be a stimulant to the better understanding of hydrogen-bonded systems.

Is the scandal, that so little is known about the interactions of macromolecules and their aqueous environment, about to be removed? There is at least a ray of hope in the work of Stephanos Pnevmatikos, from the Greek Research Centre of Crete but now working at Los Alamos, New Mexico, who has made an interesting calculation in a schematic one-dimensional structure (*Phys. Rev. Lett.* **60**, 1534; 1988).

Here is the basis for good cheer. First, the calculation is that of the dynamics of hydrogen bonding in a model structure that may have something in common with the structure of water (strictly speaking, ice). Second, it provides a good qualitative account of some known properties of the defects that arise in ice. Third, the result (with some simplifying assumptions) is a little simpler than one might expect. Fourth, on the face of things, the calculation is bound to stimulate attempts at generalization. Time will tell how many heads are broken in that process.

In outline, at least, there is general confidence in the picture of the hydrogen bond that has grown up over the past 30 years. The statics of the situation are simply described. Hydrogen atoms fit between pairs of atoms such as oxygen, nitrogen or fluorine in such a way that the energy of the system varies smoothly with the position of the proton. Both calculations and measurement show that there are two positions of the proton at which the energy is a minimum, and that these are symmetrically placed when the bonding atoms are alike. Other things being equal, the proton can be in one of two states, between which it can hop (or even reach by tunnelling), giving the distinctive energy attributable to hydrogen bonding by quantum mechanics. With only a few hydrogen bonds to be dealt with, understanding can be easily quantified.

The snags arise when the whole of a structure is held together by hydrogen bonds, as is ice, and when both the protons and the heavier atoms are free to move. While there is nothing to prevent single protons from moving from one accessible equilibrium position to another, the states so reached will be penalized in the usual Boltzmann distribution by the redistribution of electrostatic charge which they entail. Statistically, the much more common movements of the protons are inherently cooperative. To calculate the thermodynamic properties of ice requires the enumeration of all pos-

sible circuits by means of which these movements can be accomplished, which puts the problem in the same category as the solution of a three-dimensional Ising lattice.

Pnevmatikos has sensibly restrained his ambitions, at least for the time being, by quarrying a one-dimensional problem out of the structure of solid ice, and by dealing with it as if it were a problem in strictly classical dynamics. He remarks that at least some of the circuits of cooperative proton-hopping in real ice will consist of one-dimensional chains of oxygen atoms running zigzag through the lattice, with a single proton between each pair of oxygens and the other proton in each water molecule dangling out on the side. This problem, not that of ice but merely a representation of it, seems and is a little like that of *trans*-polyacetylene, again a zigzag structure in which chemical bonds can oscillate between one link in the chain and its neighbour.

In the language of billiard-ball mechanics, there is therefore a zigzag chain of heavy atoms (oxygen) which oscillate about equilibrium positions and of protons which can hop (not tunnel) between equilibrium positions which, in the real world, will themselves vary with the separation of the pair of heavy atoms enclosing them. Then there is an interaction between the two systems. Pnevmatikos's success in squeezing some exact results from these complications seems to rest to a large extent on a judicious choice of the functions used to represent the energy of the interaction between the two lattices.

The trick is to represent the potential energy of an oscillating proton by a simple function of $\cos(4\pi x/L)$, where x is the displacement of the proton and L the distance between neighbouring oxygen atoms, and to couple the two vibrating systems together by means of an energy term which is the product of the instantaneous distance between a pair of neighbouring oxygen atoms and a parameter representing the displacement of the intermediate proton from the two minima. That is where the problem becomes non-linear. There is room for endless argument about the realism of these assumptions, although charitable people will marvel that they work at all.

It will be easier to tell the importance of the other assumptions leading to exact solutions of the problem when Pnevmat-

ikos has published the promised full account of his work. (Waiting to see the details of what people have done always seems necessary when the topic is particularly interesting.) One crucial step in the argument is to arrange that the coupling between protons (only nearest neighbours are accounted for) is strong enough for the variation of their displacement along the zigzag chain to be sufficiently regular for the displacements to be represented by a smoothly varying function. So the statement of the problem boils down to a pair of partial differential equations linked in a non-linear fashion.

That, naturally enough, is not the end of the matter. In the account now published, Pnevmatikos confines himself to a search for soliton solutions — patterns of displacement propagating with constant speed along the zigzag chain, helped in part by the good luck that one of the equations is that known as the sine-Gordon equation. As it happens, two of the recognizable classes of soliton solutions represent patterns of displacement corresponding to the two known classes of dynamic defects in ice.

The defects take two forms. First, there is that in which two protons in the zigzag chain crowd in on a single oxygen in the chain, forming an H_3O^+ ion and, somewhere else, creating a vacancy or an O^- ion. Then there is the pattern in which the proton displacements are those that would arise if the orientations of neutral water molecules were changed in a pulse-like fashion so as to vary the numbers of protons between different O-O intervals along the chain. These are recognizable in real ice as, respectively, ionic defects (responsible for ionic conductivity) and Bjerrum defects.

Pnevmatikos promises that it is possible to compare the properties of these inferred patterns of displacement with experimental data for hydrogen-bonded systems. It will be interesting to learn, in due course, how great is the variety of the systems to which his solutions have some relevance. It is important that what is, after all, only the crudest model of real ice should recapture such down-to-earth properties as ionic conductivity. But it is also plain there is a long way to go before it will be possible to calculate the full thermodynamic properties of even this simple one-dimensional ice. Calculating real three-dimensional ice seems as much a nightmare as a challenge. **John Maddox**

Selective breeding

Why aren't horses faster?

William G. Hill

WE are used to new records being set in men and women's track events. In the Olympic games, for example, the time taken for men to run 1,500 metres declined by 15 seconds (or 7 per cent) from 1936 to 1984. These improvements cannot be attributed to genetic change, but to better training, health, tracks and wider screening of the population. Yet despite the efforts of breeders, the winning times of thoroughbreds in the English classic horse races have not fallen substantially over the past 50 years. The lack of improvement is disturbing because the horse-breeding industry is a large and competitive business, with much attention being paid to performance and to pedigree. If limits to performance are real, they may be relevant to other species with a shorter history of intense selection. Elsewhere in this issue, however, B. Gaffney and E.P. Cunningham (*Nature* **332**, 722–724; 1988) estimate heritabilities and rates of genetic change for a measure of racing performance, and conclude that genetic variance is still present in the thoroughbred population.

Because race times are available only for winners and in a limited number of races, Gaffney and Cunningham analyse TIMEFORM ratings of three-year old horses. A TIMEFORM rating is supposedly based only on the horse's own performance, and is expressed as the weight in pounds which a horse should carry in a free handicap race. Using several methods, Gaffney and Cunningham obtain heritability (h^2) estimates of TIMEFORM ranging from 0.39 to 0.76, the highest based on son-on-sire regression. (Heritability is essentially the proportion of the variation in a trait which is genetic.) In most cases this method is least biased by common environment, but the likely explanation is that sons of the best stallions get the best trainers. Generation intervals are long, but selection differentials on TIMEFORM ratings are high, equivalent to selecting the best 6 per cent of males and 52 per cent of females; so assuming a conservative value of 0.36 for h^2 , they predict a genetic improvement in TIMEFORM of 0.92 units per year.

The standard method now used to assess breeding values of animals and thereby estimate genetic change in populations with overlapping generations is called BLUP — best linear unbiased prediction (Henderson, C.R. *Biometrics* **31**, 423–447; 1975). As long as parameters such as heritability are known and all information on which selection decisions were based is included, unbiased esti-

mates of genetic change can be obtained. In essence, the performance of contemporary animals that have parents of different ages are compared and correction is made for selection applied. Gaffney and Cunningham use the most simple BLUP model in which only sire families are identified, yet estimate a change of 0.94 TIMEFORM units per year, almost identical to prediction.

Thus, Gaffney and Cunningham conclude there has been a genetic change over

of speed. It seems to be unlikely that the environment is worsening, for even if training methods are no better, health and nutrition have surely improved over the past 50 years and the turf of the tracks is long-established. Alternatively, Gaffney and Cunningham's estimate of genetic change is too high: even their lowest h^2 values may be biased upwards by common environment through special treatment, and the BLUP estimates depend on h^2 . Gaffney and Cunningham use $h^2 = 0.36$ and 0.48 and find little difference, but because they assume effective selection the BLUP estimates may be substantially biased upwards if the true value of h^2 is near zero.

I think resolution of this issue is of importance not just to horse breeders,

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Slow progress — today's racehorses are not appreciably faster than their forebears. The engraving "In and Out" is taken from a painting by Henry Alken, Jr.

a 25-year period equivalent to more than 20 pounds of handicap. My university library is weak on horse racing, and handicapping is obviously a complicated process. But I believe that for a 2-mile (3.2-km) race, it is 1 pound (0.45 kg) per length. For the 1.5-mile Derby, a change of 0.94 TIMEFORM units per year is equivalent to at least 0.7 lengths per year, or 17 lengths over 25 years. (I guess the length of a running horse to be about 2.5 metres, so this is equivalent to an increase in speed of about 0.1 per cent per year, clearly more than has been achieved.) So why aren't horses running any faster? These results contrast with the high rates of genetic improvement being achieved in farm livestock, which are of the order of 1 per cent per year (Smith, C. *R&D Agric.* **1**, 79–85; 1984).

Assuming the genetic change they measure is real, Gaffney and Cunningham suggest it is not inconsistent with classic winning times which refer to the extremes, not the population as a whole — but this would involve a change in the distribution

punters and bookmakers (my namesake is not a known relative). If there is real genetic change, then training methods need investigating. If there is not we need to look again at estimates of heritability and genetic progress, for Gaffney and Cunningham acknowledge that theirs is not the most complete BLUP model. More particularly, we need to explain the apparent selection limit. If all useful variation has been lost and none generated, then why? Perhaps the effective breeding population is too small? Why do breeders not cross outside the thoroughbred population? Why has the population not been divided into stayers and sprinters? What is the point of selection at all?

Breeders of species other than horses have something to learn whatever the real state of affairs. Perhaps thoroughbred horse breeders could also learn from professional geneticists, which so far they have been reluctant to do. □

William G. Hill is Professor of Animal Genetics at the University of Edinburgh, Edinburgh EH9 3JN, UK.

Plate tectonics

Putting continents asunder

Andy Saunders

WHAT causes continents to split and separate? What processes accompany such splitting? Such questions have bedevilled plate-tectonic models for two decades, not least because the ocean margins are usually buried beneath several kilometres of sediment. Seismic and deep-sea-drilling studies have shed new light on the problem, indicating that the transition zone between the continental and oceanic lithosphere is highly varied in structure and composition, and that some of these passive margins comprise large volumes of igneous rock. Two models^{1,2} have recently been proposed for the formation of these volcanically active margins. White *et al.* suggest¹ that deep-seated mantle plumes are responsible for the excess magmatism. And Mutter *et al.* propose² that vigorous mantle convection, not necessarily plume-driven, is occurring. Both studies have implications for the evolution of continent-ocean transition zones, and ultimately for the mechanisms of continental rifting.

It has been known for several years that passive ocean margins, such as those in the northern Atlantic, have a varied structure (Fig. 1). Normal margins (such as the Biscay margin) seem to have undergone crustal extension before rifting, with minimal igneous activity (Fig. 2). Others (such as the margin north-west of Britain and Scandinavia) show evidence of massive outpourings of basalt. In part, these are represented on seismic sections as seaward-dipping reflectors. Initially considered to be sedimentary layers, these reflectors are now known to be extensive basaltic sheets³. It is still debated whether these sheets were erupted onto the subsiding continental margin during the

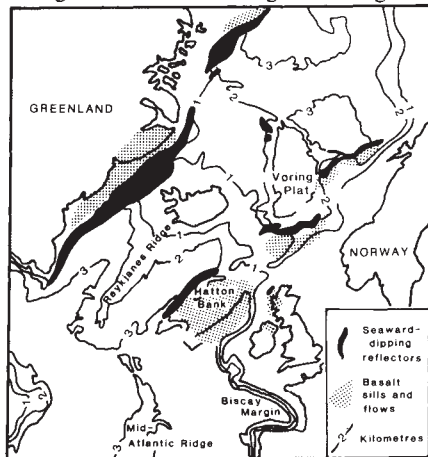


Fig. 1 Northern North Atlantic showing the distribution of early Palaeogene dipping-reflector sequences (after ref. 1). Submarine contours in kilometres.

earliest stages of rifting, or whether they represent the earliest stages of true oceanic crust, albeit considerably thickened. Recent drilling and geochemical studies^{4,5} suggest the former.

Mutter *et al.* and White *et al.* agree that the volumes of magma erupted at these margins is anomalous. White *et al.* further suggest that the transition zone along the Hatton Bank margin, for which high-resolution seismic data are available, is underplated by a thick sequence of intrusive igneous rock, suggesting that the total volume of magma emplaced along the margins of the northern North Atlantic is as great as $5-10 \times 10^6 \text{ km}^3$ —significantly greater than the Deccan flood basalt

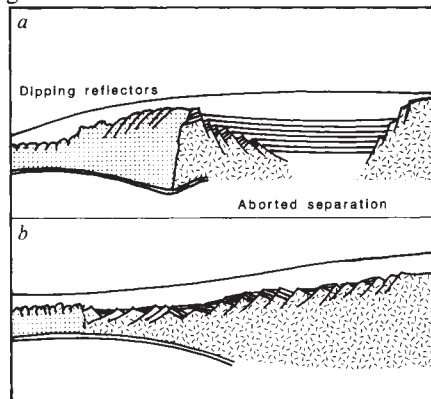


Fig. 2 The crustal structure of volcanic (a) and non-volcanic (b) margins².

province in India, for example. What causes such voluminous magmatism?

An important aspect of the thesis that Mutter *et al.* present is the sharp transition identified between oceanic and continental lithosphere along volcanic margins. This observation led them to propose that there is 'convective partial melting' beneath these margins (Fig. 3). Upwelling hot asthenospheric mantle decompresses and melts; if the upwelling is rapid, and assisted by vigorous convection, then the total volume of melt is enhanced. After the plates separate, the mantle flows into the resultant void, and the thermal contrast between the cooler lithosphere edge and the hot asthenosphere vigorously drives the convective circulation. As the continents separate, the circulation slows and is replaced by more normal passive upwelling, and thus the ocean crust produced has a more normal thickness.

What is not clear from this study is the reason for the differing style of the initial splitting of the continental lithosphere: why should the north-west margin of Britain and Norway differ from, say, the Biscay margin? Is it because of fundamental structural differences in the litho-

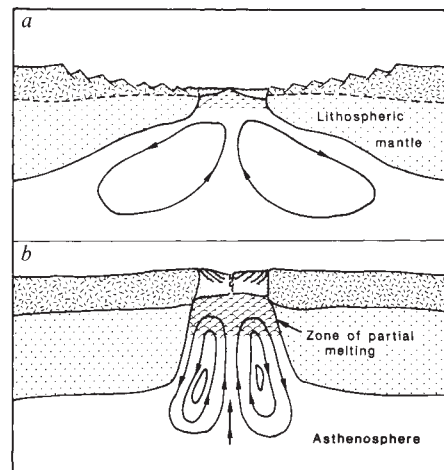


Fig. 3 Model of convective partial melting proposed² for non-volcanic margins (a) and the formation of volcanic margins (b). The sharp thermal contrast between the hot asthenosphere and the cooler lithosphere activates strong mantle convection beneath the spreading margin (arrows). Whether or not mantle conditions are propitious to the formation of such convection cells is critical to such a model.

sphere? A further problem in their model is whether or not such a small volume of mantle can convectively overturn with sufficient speed to produce the estimated volumes of magma; the mathematical constraints for the model are not detailed in this paper.

In contrast, White *et al.* emphasize the role of plumes in the evolution of volcanic margins. These are instrumental in bringing mantle with a high potential temperature (the temperature that would be reached if melting did not occur) into the melting regime. The voluminous melt production is a result of either initial heating of the continental lithosphere or direct melting of the asthenosphere.

This is an attractive model. There is evidence for plume activity in the northern North Atlantic (Iceland) and, indeed, adjacent to most other volcanic margins. The distribution of volcanically thickened margins in the North Atlantic is consistent with a mushroom-shaped plume head, about 2,000 km in diameter, beneath the originally contiguous continent. And the model circumvents the need to explain the varying style of continental rupture.

The debate could ultimately be resolved by the compositions of erupted basalts. Plume-related magmas, such as those that erupted on Iceland, have a characteristic composition in terms of trace elements (La, Nb and Ta, for example) and isotopes⁶. This signature is seen in the early eruptives on the Greenland margin, and indeed in the upper basalts from the Voring Plateau⁷. Basalts from the Rockall Plateau dipping reflectors, however, are some of the most chemically depleted basalts recovered from the oceans⁸. The Rockall data suggest that the ascending magmas interacted with continental crust

as they ascended, or that they represent an admixture of asthenosphere- and lithosphere-derived melts^{4,8}. The Vøring Plateau basalts could, on the other hand, represent uncontaminated, or less contaminated, plume-derived basalt.

Clearly, there are some important questions that could be addressed by drilling at the northern Atlantic margins. To what extent is the subcontinental mantle melted by the heating effects of an underlying plume? At what stage does the plume component begin to show through in the erupted products? What is the composition of the oceanic crust seawards of the dipping reflectors? An integrated study of rifting both before and accompanying

volcanism at a volcanic margin could address some of these problems. The best studied volcanic margins, those associated with the Icelandic plume system, should be probed by deep-sea drilling with dredging and submersible operations along the mid-Atlantic Ridge near Iceland. □

1. White, R.S. *et al.* *Nature* **330**, 439–444 (1987).
2. Mutter, J.C., Buck, W.R. & Zehnder, C.M. *J. geophys. Res.* **93**, 1031–1048 (1988).
3. Hinz, K. *Geol. Jb.* **E22**, 3–28 (1981).
4. Morton, A.C. & Taylor, P.N. *Nature* **326**, 381–383 (1987).
5. Merriman, R.J. *et al.* *Geol. Soc. spec. Publ.* **39**, 123 (1988).
6. Wood, R. *et al.* *Contrib. Mineral. Petrol.* **70**, 319–339 (1979).
7. Viereck, L.G. *et al.* *Geol. Soc. spec. Publ.* **39**, 69–83 (1988).
8. Joron, J.L. *et al.* *Init. Rep. DSDP* **81**, 783–794 (1984).

Andy Saunders is in the Department of Geology, University of Leicester, Leicester LE1 7RH, UK.

Epidemiology

Evolving control of diseases

Paul H. Harvey, Anne E. Keymer and Robert M. May

THE use and value of mathematical models for assessing the efficacy of proposed disease-control programmes have increased considerably in recent years. But many models fail to take into account significant variation in the genetics and ecology of both disease agents and their hosts, which may have important epidemiological consequences. The main problem is not a fault of the modellers, who have had little idea of the variation they are supposed to account for, but is the paucity of useful data. Technological advances are allowing the accumulation of relevant data which are being used to develop more useful models, as many of the speakers at a recent conference* demonstrated.

Karen Forsyth (University of London and Institute of Medical Research, Papua New Guinea) described both spatial and temporal variation in the genetic structure of parasite populations in a single area. Among the major diseases afflicting human populations is falciparum malaria for which an effective vaccine has still to be developed (see the recent News and Views article by Louis H. Miller, *Nature* **332**, 109; 1988). Falciparum strain-specific antigen sequences expressed in *Escherichia coli* were used to produce a simple field method that revealed the extent to which one falciparum serotype (FC27 S-antigen) is geographically and temporally localized. Although New Guinea villages separated from each other by as few as 10 kilometres have similar prevalences of the malaria (about 30 per cent of the people investigated were infected), the frequency of the FC27 S-antigen serotype differed from 0 to 42 per cent. Furthermore, at the same sites, the frequency of the serotype differs from year to year: some villages

remain virtually unaffected by the serotype, while the frequency changes in others. It is all too rarely stressed that frequency-dependent selection models for the maintenance of such genetic polymorphisms do not require constant frequencies through time. Furthermore, as Forsyth pointed out, if different serotypes have dissimilar virulence or transmission characteristics, vaccination programmes will have to be very carefully designed.

Schistosomiasis could also be controlled if suitable vaccines could be developed. Long-term studies on infected human populations are revealing important factors influencing disease transmission that may be vital in the development of vaccines (A.E. Butterworth, University of Cambridge). It is becoming increasingly clear that the lower rates of superinfection observed among older age groups results largely from acquired immunity, and less than has sometimes been suggested from age-related rates of exposure. If acquired resistance has evolved, then the prospects for control by vaccination are more hopeful. But the mechanisms of natural immunity suggest that the development of vaccines will not be straightforward. The rate at which young children develop immunity following initial infection seems to be very slow. One clue to the reasons for this is the observation that the susceptibility of young children to reinfection nine months after treatment is positively correlated with high levels of anti-egg antibodies in blood collected before treatment. Butterworth argued that individuals bearing egg-producing schistosomes produce antibodies to polysaccharides characteristic of those eggs. These same antibodies also bind to the surface of newly infecting schistosomes (schistosomula), where they are ineffective and also block the binding sites that effective

antibodies would use. Perhaps the eggs are providing protection for incoming relatives (Butterworth)? An alternative evolutionary interpretation is that the schistosomula have evolved surface coatings which mimic those of schistosome eggs.

There is a case for a more important role of exposure compared with immunity in geohelminth epidemiology (D.A.P. Bundy, Imperial College, London). Stool silica content was used in the first quantitative assessment of the extent to which observed patterns of infection intensity can be explained by the ingestion of soil contaminants. Childhood peaks of geohelminth infection may result largely from behaviourally enhanced exposure. Variation in infection among individuals, as well as age groups, is characteristic of geohelminth infection, reflecting the possible impact of many variables including exposure, immunity, physiology and nutrition. Preliminary studies in the Caribbean suggest correlates between infection intensity and HLA class II antigens, and imply a genetic role in susceptibility.

The theme of genetically determined resistance to infectious diseases in humans was developed by D. Weatherall (University of Oxford). Evidence is accumulating that polymorphisms of red cell haemoglobin, enzymes and membrane proteins have been maintained by the relative resistance of various genotypes to malaria; that blood-group heterogeneity, including secretor status, may have resulted from varying susceptibility to bacteria, virus and yeast infection; and that the HLA-DR system may be involved in modifying the clinical course of bacterial, viral and parasitic infection. Recombinant DNA technology is of increasing interest to evolutionary biologists: for example, the geographical distribution of restriction-fragment length polymorphisms suggests that the sickle-cell gene arose independently in west and east Africa. Similar evidence indicates that both the β -thalassaemia and glucose-6-phosphate dehydrogenase mutations have arisen several times in different populations, and reached their current high frequencies by selection.

Most plant viruses are transmitted between plants by specific vectors, ranging from insects and mites to nematodes and fungal spores. Using monoclonal antibodies to probe the molecular structure of some of these viruses, B.D. Harrison (Scottish Crop Research Institute, Dundee) reported a great deal of naturally occurring variation. Such variation, moreover, shows interesting patterns which invite evolutionary explanation. For instance, it appears that viruses that are transmitted 'mechanically', by simply adhering to an insect vector's proboscis, have less variable coat proteins than those that actually enter the vector as part of the transmission process; it could

* The Epidemiology and Ecology of Infectious Disease Agents Royal Society Discussion Meeting, 17–18 February 1988.

be that the greater variability of this latter group of viruses helps them deal with the cellular and humoral immune responses mounted by insect vectors.

In parallel with our increased understanding of the molecular basis of the transmission between individual hosts and pathogens or parasites is our better understanding of the overall population biology of host-parasite associations. Combining mathematical models with empirical studies of grass growth under grazing and of the epidemiology of trichostrongylid parasites of ruminants, for example, shows that these gastrointestinal nematodes can establish themselves — under some stocking regimes — at levels which result in significant depression of herbivore feeding rates (B. Grenfell, University of Sheffield). Less obviously, such nematode infections might stabilize systems that would otherwise be overgrazed to destruction, thus modifying patterns of herbivore productivity with stocking density in such a way as possibly to mask inefficiencies.

Five of the six diseases targeted by the World Health Organisation's special programme for research and training in

tropical diseases are vector-borne, and three of these involve alternative (non-human) hosts (David Rogers, University of Oxford; C. Dye, London School of Hygiene and Tropical Medicine). The population dynamics of the overall human-pathogen association can thus be rather complicated, as in the trypanosome infections *Trypanosoma* (sleeping sickness) and *Leishmania* (leishmaniasis). It is possible to have patterns of infections among humans that may be steady or fluctuating, at high or low prevalence levels, depending on the population parameters of the vector and of alternative hosts (Rogers; Dye). Amid these complexities, however, clearly formulated mathematical models can be guides to acquiring appropriate data, and tools for interpreting observed trends. Or, as Rogers and Dye believe, mathematical models should be like lamp-posts used for illumination rather than support. □

Paul H. Harvey and Anne E. Keymer are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK which Robert M. May, currently in the Biology Department at Princeton University, will join later this year.

Comet Halley

Gas and dust jets in a spin

Stephen Larson

THE unexpected discovery in 1986 of discrete jets of gas in the coma of comet Halley by A'hearn *et al.*¹ is an example of the type of contribution still possible from ground-based observation of the Solar System. As often happens, the discovery raised more questions than it answered, and many investigators are still trying to synthesize the *in situ* measurements made by the armada of Halley space probes with those made from the ground. Eventually, a picture of the comet should emerge that indicates the nature of the early solar nebula. Two reports in this issue, by Cosmovici *et al.* of CN gas jets², reported on page 705, and by Jessberger *et al.* on dust jets³, on page 691, fill in a part of this picture.

Since first recorded by Hevelius in 1682, comet Halley has become famous for its jets of dust which can make spectacular spiral patterns in its coma as the nucleus rotates (see figure). Models of dust ejected from localized areas on the comet's rotating nucleus by gas drag and perturbed by solar radiation pressure can explain the ever-changing jet configurations recorded from Earth when the comet was active in 1910 (ref. 4) and again in 1985–86. The spectacular pictures from the Vega and Giotto cameras show that most of the dust emission comes from vents that comprise less than 20 per cent of the surface area.

The rest of the surface seems to be an inactive crust of low-albedo, refractory material.

In contrast, the distribution of the gas component of the coma was thought to be quite different from that of the dust. The distribution observed in long-slit spectra and monochromatic images isolating specific spectral emission bands of observable neutral daughter species in other comets show distributions which can usually be successfully represented by models of dissociation of parent species, invisible from the ground, expanding radially from sublimating ices in, or very near, the nucleus. The rapid diffusion of daughter species far from the point of emission of the parents should produce a very uniform spatial distribution in the coma.

As part of the International Halley Watch programme, optical filters isolating the principal emissions were made available to several ground-based observers to study the distribution of gases, ions and dust. Similar, but not identical, filters were used in the Giotto multicolour camera. Gas jets were first seen in video images taken with the 61-cm reflector telescope at the Perth Observatory through the filter isolating cyanogen (CN) emission in April 1986. A'hearn and co-workers saw three pinwheel-shaped jets in the raw images displayed on the television

monitor. The contrast of the jets at that time was sufficiently high (at about 15 per cent), and the dilution by background continuum from dust was sufficiently weak at CN-emission wavelengths, for special processing to be unnecessary for their identification. Later, an azimuthal scale adjustment algorithm was applied to make the CN jets clearer and more reproducible. They were observed to distances greater than 60,000 km from the nucleus and tend to 'unwind' anticlockwise, consistent in sense with the dust jets and the projected rotation of the nucleus.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The variable structures of the coma of Comet b 1881 were sketched by E.B. Knobel¹⁰, who noted its irregular fan of jets and the eccentric position of the nucleus.

There are similar features in the C₂ band, although these are obscured by more severe continuum contamination⁵. CN jets in images taken at the Boyden Observatory in South Africa correlate poorly with the dust jets⁶, as shown in Fig. 1 of Cosmovici *et al.*² on page 706 of this issue. Images taken at the Catalina Observatory in Arizona as comet Halley approached show CN jets in January 1986, but none in November or December 1985 (ref. 6). Cosmovici *et al.*² demonstrate that CN, C₂ and C₃ are all co-spatial, implying that they have a similar origin. Two relevant characteristics of the gas jets are that they are not as curved as the dust jets, which is consistent with their higher ejection velocities, and that there are typically three jets or fewer at any one time.

The mechanism producing the gas jets is still debated. Combi⁷ suggests that if, by analogy with the dust jets, the CN parent

gas is emitted from discrete vents on the nucleus, then beyond the gas-dust decoupling region (more than 1,000 km from the nucleus) the tangential diffusion added to the radial outflow would produce gas trace jets spreading over an angle of 90° . A'hearn *et al.*¹ argue that as the observed full width at half maximum is about 45° , there must be another source of the CN parent far from the nucleus. Sekanina⁸ suggests that because the correlation between the CN and dust jets is poor, and the spacecraft observed many more small particles than expected, optically invisible grains (those smaller than the wavelength of light) could be a source of the CN parent. A'hearn *et al.*¹ go further, suggesting that the organic CHON particles discovered by spacecraft could be the source. Combi⁷ adds that for such small grains, the observed molecules must be liberated by photospattering.

Regardless of the mechanism, it is obvious that the nucleus of comet Halley is heterogeneous: the numbers of CN jets and dust jets seem uncorrelated, and some outbursts appear to be driven not by water sublimation, but by packets of more volatile materials such as CO_2 (ref. 9). As Jessberger *et al.* point out in this issue³, spacecraft data indicate that there are at least two compositionally distinct populations of dust grains. With encounter speeds of $68\text{--}70\text{ km s}^{-1}$, the spacecraft experimenters had the ionized products of high-velocity particle impacts, and little else, to study dust composition. Laboratory calibration of the instruments is impossible, so that theoretical models or assumed bulk compositions have to be used to interpret the data. Consequently, it is a giant step from measuring ion mass spectra to determining the unambiguous mineralogical composition of the dust.

Nevertheless, it seems that there are distinct populations of silicate and organic refractory grains. It is tempting to speculate that silicate grains make up most of the visible dust jets, and organic CHON particles create the gas jets. Are there pockets of different types of grains scattered about the comet's interior, or are the silicate grains mingled with the interior ices and the organic refractory grains confined to the crust as cosmic-ray-altered sublimation residue? The existing data will keep investigators busy for years. $\square$

1. A'hearn, M.F. *et al.* *Nature* **324**, 649–651 (1986).
2. Cosmovici, C.B., Schwarz, G., Ip, W.-H. & Mack, P. *Nature* **332**, 705–709 (1988).
3. Jessberger, E.K., Christoforidis, A. & Kissel, J. *Nature* **332**, 691–695 (1988).
4. Sekanina, Z. & Larson, S.M. *Astr. J.* **92**, 462–482 (1986).
5. A'hearn, M.F. *et al.* *Eur. Space Ag. spec. Publ.* **250** (1), 483–486 (1986).
6. Larson, S.M. *et al.* *Eur. Space Ag. spec. Publ.* **250** (2), 145–150 (1986).
7. Combi, M.R. *Icarus* **71**, 178–191 (1987).
8. Sekanina, Z. *Adv. Space Res.* **5** (12), 307–316 (1986).
9. Feldman, P.D. *et al.* *Nature* **324**, 433–436 (1986).
10. Knobel, E.B. *Mon. Not. R. astr. Soc.* **42**, 76–77 (1881).

Stephen Larson is in the Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA.

Oxide superconductors

Improving thin-film techniques

R.H. Koch and P. Chaudhari

HIGH-TEMPERATURE superconductors are now routinely made as thin films of high quality: transition temperatures (T_c) are often 90 K or more, well above the boiling point of liquid nitrogen (77 K), and critical current densities (which quench superconductivity) are typically above 10^6 A cm^{-2} . The screening against magnetic fields provided by films now rivals that of the best single crystals. These and other developments in the fabrication and use of thin-film superconductors were one topic of discussion at a recent meeting*.

Progress in thin-film studies of high- T_c superconductors has been accelerating perhaps because it is most likely that the first applications of the new superconductors will use thin films, which are well suited to microelectronic applications. Thin films are especially suited for measuring the 'energy-gap' magnitude and structure using tunnelling measurements, which could resolve the fundamental question pervading the conference: why do these materials superconduct?

A. Braginskii (Westinghouse) presented data indicating the advantages of rapid thermal annealing (flash annealing) over conventional oven annealing. The separation of the elemental components is minimized during rapid annealing, which will be important for limiting substrate and multilevel interdiffusion. The level of expertise in thin-film techniques was well demonstrated by the speed with which films of the new Bi–Sr–Ca–Cu–O superconducting material (see the recent

discussion in News and Views by Ted Forgan and Colin Greaves, *Nature* **332**, 14–15; 1988) were reported by many groups using several methods of fabrication. For example, K.V. Rao (Stockholm University) reported data on epitaxial films of these materials. Thin films with a composition $\text{Y}_2\text{Ba}_2\text{Cu}_3\text{O}_x$ (A. Kapitulnik, Stanford University) have a crystal structure analogous to the new bismuth-based compounds.

The existence of an energy gap Δ in the density of states, the magnitude of such a gap and its anisotropy are important for elucidating the mechanisms of superconductivity in high- T_c materials. Walking through the rows of posters at the meeting was like visiting a supermarket of physics: almost all techniques for measuring the gap have been tried. The range of values of $2\Delta/kT_c$ (k is Boltzmann's constant) obtained by tunnelling measurements on thin films varied from about 3 to 8. In a true application of statistical physics, a histogram of all the reported values (A. Barone, Università di Napoli) peaks near the value predicted by 'BCS' theory, $2\Delta/kT_c=3.5$.

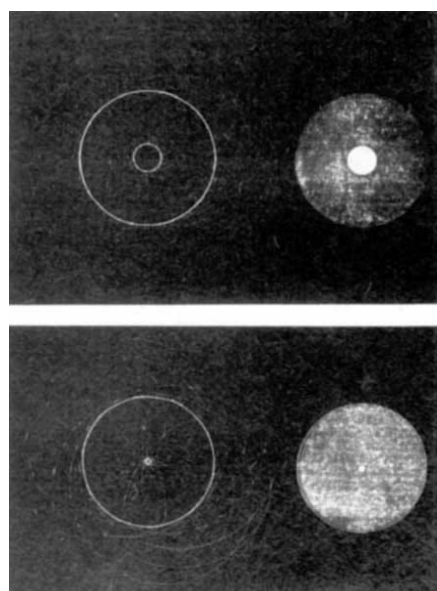
'Andreev' reflections of electrons off a silver/YBaCuO₂ interface, which occur only for electron waves (beams) with less than the band-gap energy, give $2\Delta/kT_c=3.6\pm0.6$ (P.J.M. van Bentum, University of Nijmegen). van Bentum also offered an explanation for the large values of the gaps sometimes measured using a scanning-tunnelling-microscope tip. Peaks in conductivity, like those due to tunnelling, can be seen when the applied voltage across the tunnel gap is

*International Conference on High-Temperature Superconductivity and Materials and Mechanisms of Superconductivity, 29 February to 4 March 1988, Interlaken.



Above, suggestion as to the origin of a globular nebula with a brighter central portion. The luminosity of the fainter portion is due to the collisions which occur along the sphere of intersection represented by the larger circle. After collision the meteorites will travel in new orbits, and there will be an additional sphere of intersection, represented by the smaller circle.

Below, suggestion as to the origin of a nebulous star. The orbits of the inner set of meteorites are very elliptic, so that the shell of intersection appears almost as a point. From *Nature* **37**, 587; 19 April 1888.



just sufficient to provide the energy to charge (and discharge) electrostatically any very small isolated particles in the tunnelling region.

J.S. Tsai (NEC) measured the anisotropy of the intrinsic gap of $\text{YBa}_2\text{Cu}_3\text{O}_x$ and finds that $2\Delta/kT_c$ in the c and a - b crystallographic directions (perpendicular to, and parallel with, the Cu-O planes) is approximately 4 and 6, respectively. The temperature dependence of the gap is in good agreement with the BCS prediction. Point-contact tunnel junctions seem to reveal spectral structure at energies above the gap which matches the phonon spectra obtained from neutron and Raman scattering experiments (Yu. A. Ossipyan, Institute of Solid State Physics).

Work on the use of thin films of high- T_c superconductors is dominated by the development of superconducting quantum interference devices (SQUIDs) — loops of superconductor broken by two 'normal' links (Josephson junctions) used to measure magnetic fields. Many groups reported the successful fabrication of 'bulk' SQUIDs starting from bulk polycrystalline samples of $\text{YBa}_2\text{Cu}_3\text{O}_x$. Most of these use weak links, presumably the grain boundaries in sintered samples, as the Josephson elements.

Two groups reported thin-film SQUIDs where attempts were made to localize the Josephson elements at chosen places on the SQUID loop. One (U. Kawabe, Hitachi) is a three-layer SQUID using a step edge in the substrate to form the Josephson element. The superconductivity of the $\text{YBa}_2\text{Cu}_3\text{O}_x$ deposited over a step in the substrate is weakened and thus forms a weak-link type of Josephson element. It is not clear how well this strategy works, but the SQUIDs operate up to 72 K. To be useful, the SQUID must be inductively coupled to a source of signal. Kawabe demonstrated an important first step in that direction by building a SQUID with a spiral high- T_c input coil deposited directly on the SQUID. The SQUID loop and input coil are separated by nearly a micrometre of MgO .

The other SQUID (built at IBM; R.H.K.), which works in liquid nitrogen, consists of a loop formed in an epitaxial film that was weakened by laser ablation in two selected places to form the Josephson elements. Techniques to grow polycrystalline films with very large grains have been developed (P.C.) to construct SQUIDs and other devices using a single grain boundary as the Josephson element. The current-voltage characteristics of a single grain boundary are similar to those of a point contact or microbridge.

Y. Enomoto (NTT Laboratories) discussed infrared optical detection using oxide superconducting thin films. SQUIDs and infrared detectors made from high- T_c materials are still inferior to those made from conventional materials

operating at much lower temperatures. Although the best performance will probably always be obtained at the lowest possible temperatures, there seems to be no fundamental reason why the high- T_c devices cannot be greatly improved. And the lure of high-temperature operation is overwhelming in some cases, particularly for space and military applications.

The use of superconducting films to transmit information was also discussed. The advantages of replacing present materials, aluminium and copper films, and optical fibre, with superconductors depends on the distance the information is to be transmitted. The possibility of using conventional semiconductor transistor technology cooled to liquid-nitrogen temperature connected by high-temperature superconducting wires was examined by A. Malozemoff (IBM Research Division). For the shortest distances, connecting transistors together on the same chip, superconducting wires give little advantage over copper wires cooled in liquid nitrogen, because the delay times are determined by the resistance of the transistors and the capacitance of the circuit. For chip-to-chip interconnects,

the delay time is a transmission delay and a superconductor could reduce the dispersion of the signal to allow faster operation. Together with this improvement, the shielding properties of the superconductor could allow a more compact chip-to-chip interconnect structure. Both intra- and interchip connectors look more attractive as the projected integration density increases.

For larger distances, say from computer to computer or from city to city, the comparison must be made with an optical fibre. W. Brinkman (AT&T) made such a comparison, showing that there is no significant improvement in time delay for a signal using superconductors, because both methods are limited by speed of light in the respective medium. The relevant figure of merit is the rate of information transmitted per second (bits per second) per fibre or line, and optical fibres are both generally superior to superconducting transmission lines and operate at room temperature. □

R.H. Koch and P. Chaudhari are at the IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA.

Meteorites

Studies of nucleosynthesis

Donald D. Clayton

THE identification in meteorites of anomalous isotope abundances trapped in grains before the Solar System formed has moved meteoritic science squarely into astrophysics. A significant new discovery of this type is presented on page 700 of this issue¹ by Ott *et al.*, whose data clearly show for the first time a unique isotope-isotope correlation line for xenon produced by the slow nucleosynthetic (s-) process. But the data also show variable correlation lines for the associated s-process krypton isotopes, revealing a complicated variability of the yields of the krypton that found its way into these samples. Three questions are raised by these findings: what do they mean for the s-process itself; are these samples characteristic of the bulk s-process in stars or of a special component of it; and by what historical astrophysical path has this anomalous sample been prepared and preserved for inclusion in the meteorites? Although the data address each question, they do not provide full answers.

As in other notable discoveries of this type, the sample is a carbonaceous acid-resistant residue of a carbonaceous chondrite (a primitive meteorite formed in the solar nebula), Murchison. Two of the co-authors, Yang and Epstein, had already prepared this sample and shown² (with one aliquot of it) that it is highly enriched in heavy isotopes of hydrogen

and carbon (D and ^{13}C), so there was good reason to expect interesting noble-gas contents. In the new experiment¹, Ott *et al.* burned stepwise another aliquot of that sample by exposing it to an oxygen atmosphere for 4 hours at successive temperatures of 700, 800, 900 and 1,080 °C, and thereafter pyrolysing it for 30 minutes each at 1,000 °C and at 1,800 °C. The isotope composition of the xenon and krypton released at each step is measured with a mass spectrometer. It is the correlations of the isotope changes from step to step, as revealed by the 'three-isotope plots' (Fig.1), that isolate the exotic component admixed with the normal. Briefly, the isotope contents released from samples containing two different components ('end members') are a linear combination of the two components, which can thus be found by extrapolation of the three-isotope plot.

Xenon and krypton show very different correlations in the three-isotope plots presented by Ott *et al.*. The measured points for xenon fall on a straight line, showing xenon to be a binary mixture of two components, one with isotope ratios very close to those of the Solar System and one exotic. This is the first sample in which the s-process xenon (the exotic end member) has been found in a mixture with only one other component, allowing stronger

conclusions about its composition. The solar xenon isotope composition (Fig. 2a) derives from two sources: the s-process occurring in stars, and rapid-neutron-capture (the r-process) that occurs in supernovae. According to the calculated rates of these processes^{3,4}, the latter process dominates in the normal isotopes of xenon in the Solar System. Also, the s-process stops at ¹³³Xe, which rapidly β -decays to ¹³³Cs (not shown) before further neutron capture can occur, leaving ¹³²Xe and ¹³⁰Xe the dominant isotopes of s-process xenon. The new data fully confirm that decomposition.

The s-process end member can be identified in a three-isotope plot by extrapolating the correlation lines to ¹³⁶Xe = 0, which eliminates the r-process isotopes confined to the normal end member of the mixture. Not only do the results presented by Ott *et al.* confirm the ratios of s-process xenon isotopes shown in Fig. 2a, but they show for the first time that ¹³⁴Xe = 0 when ¹³⁶Xe = 0, lending support to the belief that the 5.25-day β -decay of ¹³³Xe occurs much faster in the stellar source than does neutron capture by ¹³³Xe, which would result in ¹³⁴Xe production in the s-process. This is a significant demonstration.

One interpretation of the data (that I term 'cosmic chemical memory') is that this s-process Xe condensed within dust during its ejection from stars, before it could be mixed into the interstellar medium. Two other explanations for the origin of the exotic end member — spatial mixing inhomogeneities in the interstellar medium or the injection of s-process material into the forming solar nebula — seem less likely, because the exotic end member is so dominated by s-process material. Ott *et al.* correctly remind us, however, that although the chemical

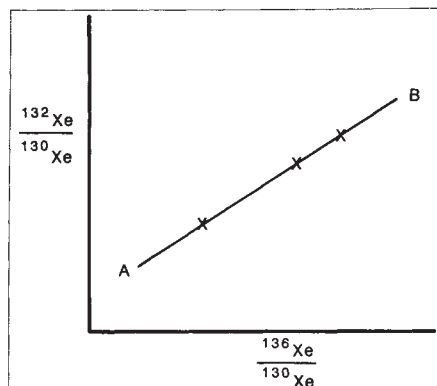


Fig. 1 A three-isotope plot (shown here for ¹³⁰Xe, ¹³²Xe and ¹³⁶Xe) can identify the parent compositions (the 'end members') of samples derived from more than one source. If three measured compositions fall on a straight line (crosses), this is good evidence that there are only two distinct components that lie on the extrapolated line, although the exact positions are indeterminate without other measurements. A mixture of three components would give data points lying in a triangle.

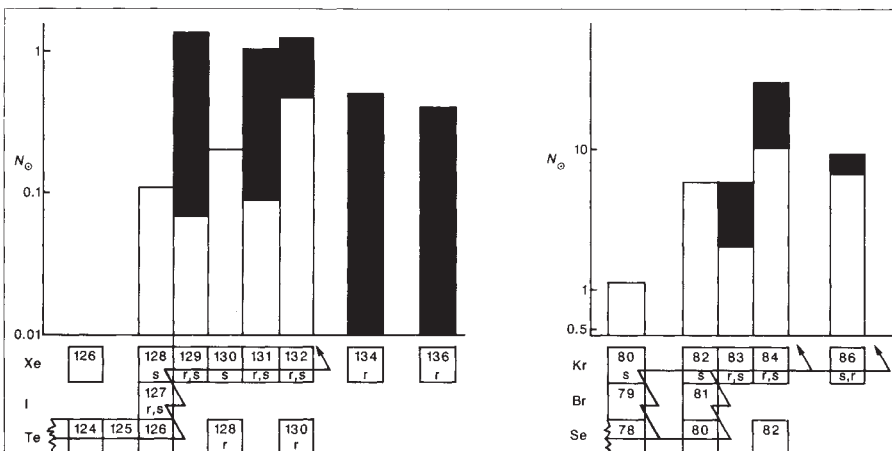


Fig. 2 Histograms of the solar abundances ($N_{\odot}$) of xenon (a) and krypton (b), and the calculated^{3,4} proportion of their s-process component (clear). The r-process component — the difference (black) — is only schematic owing to the logarithmic scale. Also shown are the relevant portions of the chart of nuclides, showing the mass number and generating process (r or s) for each isotope. The bold line shows the neutron-capture paths for the s-process. Rapid β -decay occurs at several isotopes (backward diagonals). Competition between neutron-capture and β -decay at ⁷⁹Se and ⁸⁵Kr can have important consequences for krypton isotopes, but no similar effect occurs for xenon.

carriers contain in their 900 °C fraction relatively pure s-process xenon, one cannot conclude that this is identical to bulk s-process xenon in the solar mix. The absolute amount of xenon gas measured is small enough for the carriers to have condensed in a special source that is not a dominant contributor to bulk xenon. This ambiguity plagues to some extent all of the consequences of isotopic anomalies for nucleosynthesis theory.

One might have expected that the krypton gas released from the same samples would carry an equally clear s-process signature. Ott *et al.* show that this is not the case. The nuclide chart for krypton and the predicted abundance chart show that the s-process is expected to contribute to all isotopes; however, the relative contribution to ⁸⁶Kr is uncertain because it involves competition at ⁸⁵Kr between neutron capture and β -decay, for which the half life is 10.8 years. One can state with confidence only that the s-process end member should be best indicated by the ⁸⁴Kr/⁸²Kr ratio, because ⁸⁴Kr is the only isotope in which the r-abundance is expected to exceed the s-abundance. Measured cross-sections⁶ lead one to expect ⁸⁴Kr/⁸²Kr $\approx$ 2 in s-process krypton. But ratios that low are not indicated by the correlation lines. Indeed, Ott *et al.* show that the extrapolation to ¹³⁶Xe = 0, which should give the pure s-component, intercepts at ⁸⁴Kr/⁸²Kr = 2.87. This is a small, but nonetheless significant, discrepancy from s-process theory, and could indicate that the s-process source being sampled is unusual, and not the major bulk source. The good three-isotope correlation line involving the three uncomplicated isotopes (^{82,83,84}Kr) is consistent with the anticipated s-process ratio ⁸³Kr/⁸²Kr = 0.35 and the surprising ratio ⁸⁴Kr/⁸²Kr = 2.87 indicated above.

The other two isotopes, ⁸⁰Kr and ⁸⁶Kr, cannot be correlated at all, which is another surprise. It is as if the s-process ratios ⁸⁰Kr/⁸²Kr and ⁸⁶Kr/⁸²Kr differ from one temperature of release to another. This means either that the exotic krypton is not s-process krypton, even though correlated xenon is, or that the stellar carriers have recorded yields from differing stellar circumstances. The latter is not unlikely, because the yields of both ⁸⁰Kr and ⁸⁶Kr depend upon the detailed competitions between neutron capture and β -decay at ⁷⁹Se and ⁸⁵Kr respectively (see Fig. 2b). Furthermore, because ⁸⁰Kr and ⁸²Kr are solely s-process isotopes (shielded from r-process production), the variability in their ratio is itself an indication of differing s-process components. The β -decay of ⁷⁹Se, which is needed for ⁸⁰Kr production, is very temperature sensitive, and both branch points depend upon the neutron density.

Ott *et al.* suggest that the different oxidation-temperature fractions are releasing krypton from carriers formed from differing nucleosynthesis products. It will be very hard to ascertain whether these are distinct giant stars, having differing s-process conditions internally and differing isotopic compositions in their atmosphere, or whether it is a time-dependent change of those circumstances in one class of star. These data issue a strong challenge to nucleosynthesis theory. □

- Ott, U., Begemann, F., Yang, J. & Epstein, S. *Nature* **332**, 700–702 (1988).
- Yang, J. & Epstein, S. *Nature* **311**, 544–547 (1984).
- Clayton, D.D. & Ward, R.A. *Astrophys. J.* **244**, 1000–1006 (1978).
- Beer, H., Käppeler, F., Reffo, G. & Venturi, G. *Astrophys. Space Sci.* **97**, 95–119 (1983).
- Clayton, D.D. *Q. Jl R. astr. Soc.* **23**, 174–212 (1982).
- Walter, G., Beer, H., Käppeler, F., Reffo, G. & Fabbri, F. *Astr. Astrophys.* **167**, 186–199 (1986).

Donald D. Clayton is in the Department of Space Physics and Astronomy, Rice University, Houston, Texas 77251, USA.

Molecular evolution

DNA-based phylogenies of the three chimpanzees

Jared M. Diamond

SOME of the most acrimonious disputes in biology today are those among advocates of different taxonomic methods, closely rivalled by those among students of human origins. Recent studies of human origins by two taxonomic methods (DNA sequencing and DNA/DNA hybridization) are helping to resolve both disputes, while making clear how much the average rate of DNA evolution varies among phylogenetic groups^{1,2}.

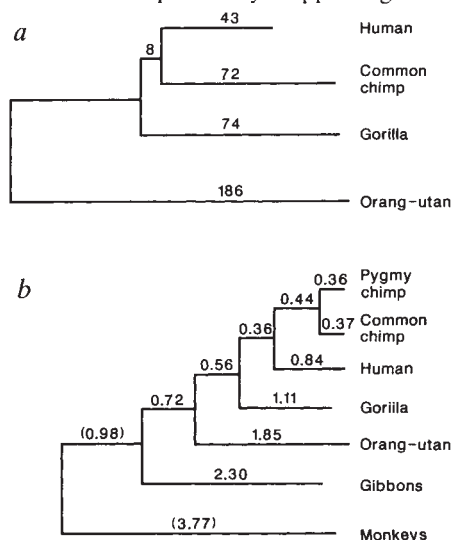
Taxonomy, which formerly relied mainly on anatomical characters, was enriched in the 1950s by sequences or electrophoretic mobilities of proteins used as molecular characters. For taxonomists, a protein sequence has two advantages over an anatomical trait: it is specified by one genetic locus rather than by many; and it is subject to little or no non-genetic modification. But a protein retains some disadvantages of an anatomical trait, such as expressing only a very small portion of the genome and being subject to natural selection and convergence.

The interest of taxonomists thus shifted in the 1970s to DNA itself, particularly to non-coding DNA, studied either by sequencing, to identify and count up interspecific differences in nucleotide sequences of a fraction of the genome; or by single-copy DNA/DNA hybridization, to measure directly the summed interspecific differences in sequence for the whole genome (most of it non-coding) by the melting-point depression of reassociated DNA strands from the two species. Application of the latter method³ in 1984 sparked the controversy that led to the two recent experimental papers^{1,2}.

Miyamoto *et al.*¹ apply the sequencing approach to the longest continuous stretch of orthologous DNA yet determined for humans and three of the four great apes, a 7.1-kilobase segment of the η -globin gene region. Sibley and Ahlquist² describe the application of DNA/DNA hybridization to humans, all four great apes (including pygmy chimpanzee, omitted in the sequencing study), two gibbon species and seven cercopithecoid (Old World monkey) species. As shown in the figure, the most parsimonious deduced phylogenies from these studies yield the same branching patterns as each other and as the 1984 study. Three conclusions follow.

First, although virtually all recent authors agree that the gibbons and then the orang-utan were the first apes to branch off the hominoid ancestral line, there has been no consensus about the

branching sequence of the gorilla, the two chimpanzee species (pygmy chimpanzee and common chimpanzee) and humans. Most anatomists contrast humans with the other three species lumped as 'African great apes' (the gorilla and the two chimpanzees), implying that these three African apes are closer to each other than any of the three is to humans. But the obvious characters superficially supporting this



Most parsimonious hominoid phylogenies deduced from sequencing of the η -globin gene region¹ (a) and from single-copy DNA/DNA hybridization of the whole genome² (b). Numbers represent the number of base substitutions plus gap events (a) or the depression of mixed melting-point in °C (b) apportioned to that branch. Monkey, cercopithecoids.

view, such as humans' bipedal gait, large brain and advanced use of tools, actually provide no supporting evidence, as they did not appear until the australopithecine stage or later, long after humans had branched off from chimpanzees and gorillas. Both the sequencing and hybridization studies show that chimpanzees and humans are slightly closer to each other than either is to gorillas. Thus, the gorilla line branched off first (an estimated 7.7–11.0 Myr, or million years, ago³), followed by the bifurcation between the human line and the two 'other' chimpanzees (5.5–7.7 Myr ago³), and finally by the bifurcation between the common and pygmy chimpanzees (2.4–3.4 Myr ago³). Humans should not therefore be contrasted with African great apes, but gorillas should be contrasted with the three chimpanzees (human, common and pygmy chimpanzees).

Second, the constancy of molecular clocks has provoked another vigorous,

longstanding debate. Just as earlier conclusions of constant-rate protein clocks had to be abandoned, so too the recent results refute an earlier conclusion³ that the average rate of evolutionary change in DNA sequences is uniform among species. Instead, the average rate is faster in rodents than in hominoids and faster in fruitflies than in songbirds^{2,4,9}. For the η -globin region, the rate appears 30–40 per cent slower for the human lineage than for the lineage of the other chimpanzees or of gorillas¹, but rates averaged over the whole genome do not show obvious differences among these lineages². One main unresolved problem is whether these species differences in rate of DNA evolution are caused by differences in age at first breeding, efficiency of DNA repair, germline cell cycles or another factor^{2,4,6,7}.

Finally, the agreement between the sequencing and hybridization approaches to DNA-based taxonomy of hominoids reinforces earlier evidence that the two methods converge on the same result⁴. A moment's reflection shows why this conclusion was inevitable. In the sequencing method, homologous DNA sequences from two species are determined, the base sequences are aligned according to the best match, and the mismatches are summed to arrive at the minimum number of mismatched sites between the sequenced fragments of DNA. In the single-copy hybridization approach, the sequences of the two species align themselves by AT and CG hydrogen-bonding, and the mixed melting-point depression is considered as proportional to the number of mismatches for the whole genome. Because the sequencing method uses only a fragment of the genome, it must tend to yield errors resulting from the small number of mismatches and from idiosyncracies of individual genes, but must converge on the results of the hybridization method as the size of sequenced fragment increases.

Despite this compelling empirical and theoretical evidence for convergence of the two methods, some adherents of the taxonomic philosophy called cladism criticize the hybridization approach while accepting the sequencing approach^{10–14}. Although the critics' objections had been refuted previously^{2,7,15}, the new results^{1,2} help illustrate the flaw in the oft-repeated objection that DNA/DNA hybridization fails to distinguish primitive from shared derived traits, a *sine qua non* of cladistic procedures. In fact, it is not generally appreciated that neither DNA sequences, morphological characters nor any other characters permit this distinction unequivocally for individual traits. If species A and B but not C share a character, A and B might have evolved it separately, or on the other hand they might both have inherited it as a primitive character from an ancestor while C lost it. If fossils show that the

ancestor lacked it, A and B could have inherited it as a derived character from an intermediate ancestor or else could still have evolved it separately.

To appreciate that these are not just far-fetched theoretical alternatives, consider the two shared characters (knuckle-walking and thin dental enamel) that Martin¹⁰ and Andrews¹² cite as (the sole) apparently derived morphological traits linking chimpanzees and gorillas and separating them from humans. Other anthropologists⁶, in disagreement with Martin and Andrews, suggest that proto-humans also practised knuckle-walking and then lost it, or that chimpanzees and gorillas evolved knuckle-walking (and thin enamel) independently. The same ambiguities apply to DNA sequences. For instance, the presence at η -globin position 5,153 of cytosine in chimpanzee and gorilla contrasting with adenine in human and orang-utan nominally suggests cytosine to be a shared derived trait linking chimpanzee and gorilla. But Miyamoto *et al.*¹ implicitly interpret it either as lost by proto-humans or else evolved independently by chimpanzee and gorilla. The η -globin sequences include three such features construed as pseudo-synapomorphies (at positions 5,153, 5,156 and 6,808) shared by chimpanzee and gorilla but not human or orang-utan — nearly half as many as the eight features interpreted as true synapomorphies (shared by chimpanzee and human but not gorilla or orang-utan).

The sole solution to this dilemma is to analyse a very large number of characters, as DNA/DNA hybridization in effect does (for example, by estimating about 32 million base-pair mismatches between humans and the 'other' two chimpanzees). Although interpretation of any single DNA sequence shared between A and B but not C or an ancestor is equivocal, a finding that A and B share many such non-coding sequences would identify common descent rather than independent evolution as the correct explanation.

Is sequencing or DNA/DNA hybridization the more useful technique? The latter has disadvantages for studying very distant taxa⁹. Neither is ideal for analysing relations of conspecific races or of very closely related species, which may differ too little in the nuclear DNA (but see refs 8, 9). Mitochondrial DNA, which evolves more rapidly than nuclear DNA, is better for such studies. But if one's interest is in the phylogeny of well-differentiated, not-too-distant species, DNA/DNA hybridization has two decisive advantages.

First, it measures the net sequence divergence between whole genomes, but sequencing can currently deal with only tiny fractions of genomes (about 0.0004 per cent in the study by Miyamoto *et al.*), and those fractions can vary in their evolution. Second, it offers higher accuracy and statistical power than is feasible with

sequencing. Felsenstein¹⁷ discusses how to calculate the number of base pairs that must be sequenced to yield the same statistical power as a single DNA/DNA hybridization value, given the standard error of the latter. If one considers that newer techniques to sharpen DNA/DNA hybridization curves reduce the standard error^{8,9} and that hybridization values are routinely measured in replicate (for example, 64 replicates for human $\times$ common chimpanzee in ref. 2), more than 100,000 orthologous base pairs of human and common chimpanzee would have to be sequenced to achieve a similarly low standard error. Although it may some day be feasible to sequence so much material from two species, sequencing will remain a brute-force, indirect method to obtain a measure of total mismatches that DNA hybridization yields directly.

All this is not to say that measuring DNA sequences and other taxonomic characters is a waste of time. Sequencing is obviously essential if one is interested in the evolution of the particular gene sequenced, rather than in taxonomic relationships. The DNA/DNA hybridization technique increases rather than decreases the evolutionary interest of protein sequences and morphological traits, which formerly had to be used in circular fashion to infer both overall taxonomic relationship and the evolution of the studied trait. DNA/DNA hybridization can now be used to establish taxonomic relationships independently, so that evolution of proteins or anatomy can be established unequivocally. However, for the purpose of deducing phylogenies I expect that DNA/DNA hybridization will be the method of choice in future. Thus, I see the study by Miyamoto *et al.* not as the harbinger of even larger-scale brute-force sequencing efforts by taxonomists, but instead as resolving any lingering doubts that such efforts confirm the results of hybridization methods. □

1. Miyamoto, M.M. *et al.* *Science* **238**, 369–373 (1987).
2. Sibley, C.G. & Ahlquist, J.E. *J. molec. Evol.* **26**, 99 (1987).
3. Sibley, C.G. & Ahlquist, J.E. *J. molec. Evol.* **20**, 2 (1984).
4. Britten, R.J. *Science* **231**, 1393–1398 (1986).
5. Wu, C.-I. & Li, W.-H. *Proc. natn. Acad. Sci. USA* **82**, 1741–1745 (1985).
6. Catzeflis, F.M. *et al.* *Molec. biol. Evol.* **4**, 242–253 (1987).
7. Sibley, C.G. *et al.* in *Evolutionary Biology* Vol. 21 (eds Hecht, M.K. *et al.*) 97–125 (Plenum, New York, 1987).
8. Hunt, J.A. *et al.* *J. molec. Evol.* **17**, 361–367 (1981).
9. Caccone, A. *et al.* *Molec. biol. Evol.* **44**, 343–350 (1987).
10. Martin, L. in *Major Topics in Primate and Human Evolution* (eds Wood, B. *et al.*) 161–187 (Cambridge University Press, 1986).
11. Andrews, P. in *Major Topics in Primate and Human Evolution* (eds Wood, B. *et al.*) 107–129 (Cambridge University Press, 1986).
12. Andrews, P. in *Molecules and Morphology in Evolution: Conflict or Compromise?* (ed. Patterson, C.) 23–53 (Cambridge University Press, 1987).
13. Cracraft, J. in *Evolutionary Biology* Vol. 21 (eds Hecht, M.K. *et al.*) 47–96 (Plenum, New York, 1987).
14. Farris, J.S. *Cladistics* **1**, 67–85 (1985).
15. Felsenstein, J. *Cladistics* **2**, 130–143 (1986).
16. Pilbeam, D. *Am. Anthropol.* **88**, 295–312 (1986).
17. Felsenstein, J. *J. molec. Evol.* **26**, 123–131 (1987).

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Daedalus

Bright noise

THE thermograph, which images a scene in surface temperature, has many medical and industrial uses. Daedalus is working on an instrument to image, not the heat given out by every point of a scene, but the sound. In effect it produces a sound-map of the scene.

Rather elegantly, the 'sonograph' will image the whole scene simultaneously in sound and vision. It has a lens with the same focal length for sound as for light. Daedalus calculates that such a lens can be made as a doublet, one element of which is the liquid perfluoroheptane, while the other is a mixture of 80 per cent helium and 20 per cent air. The liquid and gas elements are separated by a curved transparent membrane, its radius determined by the internal pressure; the front and rear faces are conveniently plane. For good focusing, its diameter should be at least a wavelength of the lowest sound to be imaged — more than one metre for middle C! The practicalities may dictate something smaller, with some sacrifice of bass resolution. The whole device will resemble a view-camera, with a ground-glass screen that the observer scrutinizes from under a black cloth.

To display the sonic image superimposed on the optical one, Daedalus will exploit the poetic phenomenon of 'the burning sea'. This luminosity of disturbed tropical sea water results from bioluminescent marine algae, like the dinoflagellate *Noctiluca miliaris*, which flashes when mechanically stimulated. An agar culture of *Noctiluca* on the focusing screen of the sonograph will flash brightly at those points of the image that are emitting sound. A stethoscope touched to such a point will reveal what the sound is.

Engineers should rush to buy sonographs. All the folklore of machine diagnosis by sound — the elusive rattle of loose parts, the squeak of dry bearings, the grinding of worn gears — will be reinforced by precise location of the noise. Silencing and sound-proofing will make giant strides: the sonograph will not only identify the real problem, but will show immediately the effectiveness of counter measures. The hi-fi fanatic will at last be able to image the 'sound stage' of his stereo equipment, and modify speaker positions and so on until his stethoscope reports every instrument of the orchestra in its proper place.

The sonograph will also have an impact on politics. Conference stewards will be able to identify distant hecklers, or insufficiently hysterical contributors to a standing ovation. In the televising of parliaments, it would usefully locate the sources of muttered backbench dissent. David Jones

The rocky road to panspermia

SIR—Recent evidence that the SNC meteorites originated on Mars raises the question of whether large impacts on Earth may eject rocks that could fall on Mars (or other planets in the Solar System) and, if so, whether they might contain spores or some sort of viable microorganisms that would have the opportunity to colonize Mars.

To answer the first part of the question, I use the theory of spall ejection^{1,2} to examine the mass and velocity of material ejected from the near vicinity of an impact. Because the spall mechanism only operates near a free surface, ejected rocks must originate close to the surface, consistent with data from both the martian and lunar meteorites. The near surface is also the most biologically active region of the Earth's crust, so the chance of ejecting spores and microorganisms with the spalled material is relatively high. The theory³ indicates that the velocity of ejection v_{ej} is a function of the projectile impact velocity U , its radius a , the effective depth of energy deposition d ($d \approx a$ for most impacts) and the distance along the surface s from the point of impact to the position of the ejected material

$$v_{ej} = \frac{Ud}{a} \left(\frac{a}{r} \right)^{2.87}$$

where $r = \sqrt{s^2 + d^2}$ is the distance from the effective centre of energy deposition to the surface ejection point. The ejection velocity is only a weak function of depth of origin near the surface.

One cannot neglect the effect of atmospheric interference. Numerical computations of the effect of a 10-km diameter asteroid on the Earth's atmosphere⁴ indicate that such a projectile clears the atmosphere away from its immediate vicinity, leaving behind a 'hole' that takes minutes to fill and through which early, fast, spalled ejecta may escape (the transit time of ejecta travelling at more than 11 km s^{-1} through 100 km of atmosphere is only of the order of 10 seconds). To be conserva-

tive, however, I performed spall computations in which I assume that the atmosphere is present at its normal density. In a large impact, the sheer mass of high velocity spalled material is so large that it exceeds the mass present in the atmospheric column, so the spalled surface rocks may not be significantly slowed by the atmosphere.

For the purposes of the computation, I average the momentum of the spalled material that was ejected faster than the Earth's escape velocity and the atmospheric gases along the escape trajectory. If the final velocity of both the spalled material and the atmosphere still exceeded escape velocity, then this material is counted as having escaped. If the average is less than escape velocity, then it is presumed that the material is stopped by the atmosphere and falls back to Earth. This criterion implies that the initial ejection velocity must be larger than the Earth's escape velocity. It is easy to show that v_{ej} must exceed

$$V_{lim} = \frac{V_{esc}}{1 - \frac{m_{air} \beta c_L V_{esc}}{2 P_{max} a}}$$

where m_{air} is the mass of air per unit area along the escape trajectory. For the computations performed here, it was taken to be 10^4 kg m^{-2} , assuming a nearly vertical trajectory, as appropriate for the fastest ejecta². $\beta \approx 4$ is the ratio between the duration of the trailing and rising portions of the impact generated stress pulse² and c_L is the speed of the compressional stress wave in the target rocks. P_b is the maximum pressure experienced by the material at the base of the spalled surface: it controls how much spalled mass is available to push away the atmosphere. I have taken it here to be the pressure at the onset of melting, about 70 GPa.

Much of the spalled material, while protected from the maximum pressure occurring in the shock wave by rarefactions from the free surface, nonetheless suffers

significant shock damage. Although the peak pressure experienced by a portion of the ejected material, P_{max} , is zero at the free surface, it increases nearly linearly with increasing depth for shallow depths. The depth z_{ej} to material that has experienced pressure P_{max} is a function of the ejection velocity and material parameters given by

$$z_{ej} = \frac{2 P_{max} a}{\rho \beta c_L v_{ej}}$$

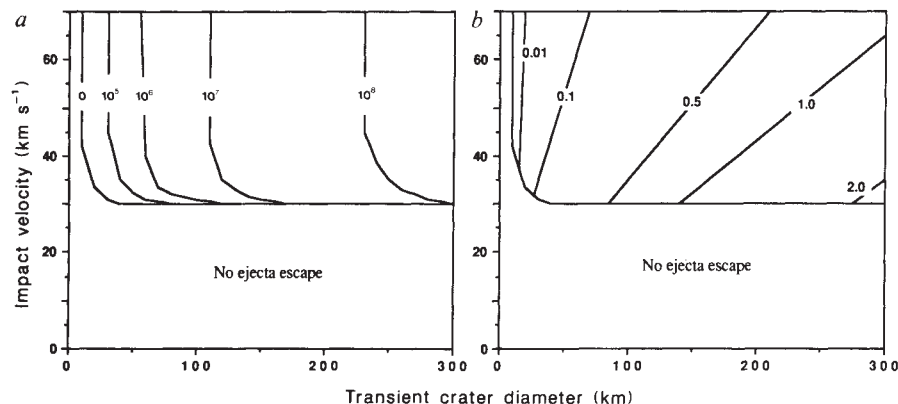
where ρ is the target's density (assumed to equal that of the projectile).

One of the SNC meteorites (EETA 79001) contains impact melt glass and may have experienced pressures as great as 50 GPa⁵. Any organisms living in the rock would certainly have been killed by the heat generated from the shock event. Only the rocks nearest the surface that receive the maximum protection from shock would escape heating to less than 100°C and thus retain viable microorganisms and spores. I take the corresponding shock pressure as equal to the stress necessary to crush the rock (when internal pores collapse, the PdV work done during the collapse greatly enhances the energy deposition, and at this point residual temperatures greatly exceed 100°C), or about 0.1 GPa (1 kilobar). The volume of ejected material thus escaping sterilization is considerably less than the total volume of spalled rock, but this near-surface zone begins with a full complement of soil microorganisms, some of which appear likely to survive the ejection process. The thickness of this zone, which I call the fecund zone, is easily estimated to be

$$z_{fecund} = \left(\frac{P_{sterile}}{P_{max}} \right) z_{ej}$$

Behind the fecund zone is a larger mass of hot sterilized rock that pushes boulders of more than 1 m diameter from the fecund zone through the atmosphere. Although some brief heating may occur at the surface of the fecund zone as the atmosphere is compressed in front of it, this heating is of short duration, as at 11 km s^{-1} the ejecta clear the lower atmosphere in a matter of seconds. The thermal penetration depth of such a heat pulse is only a few millimetres, so that most of the ejected organisms will survive.

The figure (a) shows the volume of fecund material (which is only a few per cent of the total spalled material) ejected from the Earth as a function of impact velocity and transient crater diameter (the final crater will be about 65 % larger than the transient crater size). Note that no ejecta escape unless the impact velocity exceeds about 30 km s^{-1} . This is a conservative limit: the spall model used to construct this plot assumes that no spall material is ejected faster than about 35% of the impact velocity. Direct observations



a, Volume (in m^3 marked on the curves) of fecund ejecta and b, fragment size (in m marked on curves) as a function of impact velocity and transient crater diameter. Crater diameters estimated from projectile radius and velocity using the revised Schmidt-Holsapple scaling relation.^{1,2}

on small-scale impacts⁶ indicate that some material may be spalled at speeds approaching 85 % of the impact velocity. It is clear that a 50–75-km-diameter transient crater (which would produce a final crater comparable in diameter to the observed large craters on earth) could eject roughly a million cubic meters of boulders of more than one meter diameter from the fecund zone. The diameter of the boulders is estimated by the fragmentation theory of Grady and Kipp (discussed in refs 1–3), which predicts the fragment diameter l_{GK}

$$l_{GK} = \frac{2Ta}{\rho v_{ej}^{2/3} U^{4/3}}$$

where T is the tensile strength of the rock (taken here to be 0.1 GPa). The fragment diameter estimated by this equation is shown in the figure (b).

It seems likely that the impacts that produced the craters on Earth that are greater than 100 km in diameter⁷ would have ejected millions of tons of near-surface rocks carrying viable microorganisms into interplanetary space, much in the form of boulders large enough to shield those organisms from ultraviolet radiation, low-energy solar cosmic rays, and even galactic cosmic rays. Under such circumstances spores might remain viable for long periods of time. Even microorganisms active at the time of ejection might have been preserved for considerable periods by lopholization in hard vacuum. Viable organisms preserved by this process were recovered from the Surveyor 3 camera assembly after three years on the lunar surface⁸.

No computations have yet been performed to indicate how long Earth ejecta would take to reach Mars. Similar computations on the migration of Mars ejecta to Earth⁹ indicate mean transit times of millions of years; the probability of a direct orbit is exceedingly low. Once at Mars, boulders falling to the surface may be slowed by the thin atmosphere and fragmented by aerodynamic stresses at low altitudes, exposing their interiors. This process is most important for objects of less than a meter diameter^{10,11}.

Although terrestrial organisms in these rocks would have the opportunity to colonize the planet, it seems unlikely they would find conditions suitable for propa-

gation. Planets of the Solar System should therefore not be thought of as biologically isolated: from time to time large impacts may inoculate Mars and the other planets of the inner Solar System with a sample of terrestrial life.

H. J. MELOSH

Lunar and Planetary Laboratory,
University of Arizona,
Tucson, Arizona 85721, USA

Leukocyte L1 protein and the cystic fibrosis antigen

SIR—Since its initial description as a major leukocyte protein¹, L1 has been characterized and its expression localized. Briefly, it is a calcium-binding, heat-stable multi-chain protein, with a relative molecular mass (M_r) of about 36.5 (ref. 2). Increased L1 plasma concentration is found in many infectious and malignant conditions^{3,4}, and L1 is present in leukocytes⁵, macrophages⁶ and epithelial cells⁷.

Two-dimensional electrophoresis of purified L1 shows that it corresponds to the leukocyte polypeptides RA 6–7 which are increased in rheumatoid arthritis⁸. Purified L1 from human granulocytes separates into three polypeptide chains, on two-dimensional isoelectric focusing/SDS gel electrophoresis; the M_r and pI of polypeptides I, II and III are M_r 12.5, 13.3, 8.3; pI 5.5, 6.0, 7.0 (H.B.B. and M.K.F., unpublished). Polypeptides I and II cannot be distinguished by rabbit or rat polyclonal or mouse monoclonal antibodies. The amino-acid compositions of polypeptides I and II are similar⁹, and their small differences in pI and mass probably result from post-translational modifications and/or deamidation. Polypeptide chains III and I/II are called the L1 light and heavy chain, respectively.

Odink *et al.*¹⁰ recently described two macrophage proteins, MRP-8 and MRP-14, the former of which is nearly identical to the cystic fibrosis (CF) antigen¹¹. The sequences deduced from cDNA encoding the MRP-8 and MRP-14 polypeptides¹⁰ are identical to the partial amino-acid sequences of L1 light and L1 heavy chains, respectively⁹. Post-translational modifications are, however, observed in L1 light-chain positions 1 and 6. Sequences of tryptic peptides derived from the C-terminal region of the L1 light chain all agree with the MRP-8 cDNA sequence from position 292 but no peptides corresponding to the deduced CF antigen C-terminal sequence¹¹ were detected. Like Odink *et al.*, we suggest there is a sequencing error from nucleotide 293 of the CF antigen.

Overall, there can be no doubt that the MRP-8/CF antigen and MRP-14 are identical to the L1 light and heavy chains, respectively. Note that, like the L1 antigen⁷, the CF antigen is present in squamous

mucosal epithelia¹¹. Until an appropriate name can be coined on the basis of a putative biological function, we propose that the terms L1 antigen, L1 light chain and L1 heavy chain should be used.

KRISTIN B. ANDERSSON
KNUT SLETTEN

Department of Biochemistry,
University of Oslo, Norway

HILDE BERNER BERNTZEN
MAGNE K. FAGERHOL

Blood Bank and Department of
Immunology,
Ullevål Hospital,
Oslo, Norway

INGE DALE

PER BRANDTZAEG

Institute of Pathology,
Rikshospitalet, Oslo, Norway

EGIL JELLUM

Institute of Clinical Biochemistry,
Rikshospitalet, Oslo, Norway

1. Fagerhol, M.K., Dale, I. & Andersson, T. *Scand. J. Haemat.* **24**, 393–398 (1980).
2. Dale, I., Fagerhol, M.K. & Naesgaard, I. *Eur. J. Biochem.* **134**, 1–6 (1983).
3. Sander, J., Fagerhol, M.K., Bakken, J.S. & Dale, I. *Scand. J. clin. lab. Invest.* **44**, 357–362 (1984).
4. Stockley, R.A., Dale, I., Hill, S.L. & Fagerhol, M.K. *Scand. J. clin. lab. Invest.* **44**, 629–634 (1984).
5. Dale, I., Brandtzaeg, P., Fagerhol, M.K. & Scott, H. *Am. J. clin. Path.* **84**, 24–34 (1987).
6. Brandtzaeg, P., Dale, I. & Fagerhol, M.K. *Am. J. clin. Path.* **87**, 681–699 (1987).
7. Brandtzaeg, P., Dale, I. & Fagerhol, M.K. *Am. J. clin. Path.* **87**, 700–707 (1987).
8. Willard, K.E., Thorsrud, A.K., Munthe, E. & Jellum, E. *Clin. Chem.* **28**, 1067–1073 (1982).
9. Andersson, K.B. *et al.* *Scand. J. Immun.* (in the press).
10. Odink, K. *et al.* *Nature* **330**, 80–82 (1987).
11. Dorin, J.R. *et al.* *Nature* **326**, 614–617 (1987).

Return of the 'last ribo-organism'

SIR—In accusing us of a profound misunderstanding, Maizels and Weiner¹ continue to illustrate the problems encountered as molecular biologists attempt to model the 'RNA world'.

For example, they write that cofactors containing fragments of RNA would emerge in a world of protein catalysts because "nucleotides are far better catalysts than proteins for many reactions". This is wrong. The nucleotide fragments of RNA cofactors are incidental to their chemical reactivities². In the RNA world model, their existence is a 'vestige' of early RNA metabolism³. Cofactors that emerged after the breakthrough to the 'protein world', such as biotin⁴, are not ubiquitous, lack RNA fragments, and have reactivities reflecting the improved catalytic power of proteins^{2,4}.

Indeed, Maizels and Weiner's writing is generally confused about vestigiality. If an RNA unit performs a selectable function in the modern world that is intrinsic to the chemistry of RNA and could not be performed as well by proteins, the function is probably not a vestige of the RNA world but has arisen more recently. Therefore, the modern function and intrinsic chemistry of introns and tRNA-like structures

1. Melosh, H. J. *Geology* **13**, 144–148 (1985).
2. Vickery, A. M. & Melosh, H. J. *Science* **237**, 738–743 (1987).
3. Melosh, H. J. *Icarus* **59**, 234–260 (1984).
4. Roddy, D. J. *et al.* *Int. J. Impact Eng.* **5**, 525–541 (1987).
5. McSween, H. Y. Jr. *Rev. Geophys.* **23**, 391–416 (1985).
6. Curran, D. R. *et al.* in *Impact and Explosion Cratering* (eds Roddy, D. J. *et al.*) 1057–1087 (Pergamon, New York, 1977).
7. Grieve, R. A. F. *Ann. Rev. Earth planet. Sci.* **15**, 245–270 (1987).
8. Mitchell, F. J. & Ellis, W. L. in *Analysis of Surveyor 3 Material and Photographs Returned by Apollo 12* 239–251 (NASA, Washington, DC, 1972).
9. Wetherill, G. W. *Meteoritics* **19**, 1–13 (1984).
10. Passey, Q. R. & Melosh, H. J. *Icarus* **42**, 211–233 (1980).
11. Melosh, H. J. in *Multi-Ring Basins* (eds Schultz, P. H. & Merrill, R. B. *Proc. Lunar Planet.* **12A**, 29–35 (1981).
12. Schmidt, R. M. & Housen, K. R. *Int. J. Impact Engng.* **5**, 543–560 (1987).

at the 3'-terminus of some RNA viruses is evidence against their antiquity, not for it, as Maizels and Weiner write^{1,5}.

Further, the poly(C)-forming activity of self-splicing introns cannot be viewed as a vestige of "an early RNA replicase"⁵. Drift would have destroyed this activity immediately after it ceased to be functional were it not intrinsically associated with splicing chemistry. And if poly(C)-forming activity is intrinsically associated with splicing, it would be found in self-splicing introns whenever they emerged.

We are agnostics concerning the interesting 'genomic tag' model. But it is weak because it takes an isolated aspect of viral biochemistry and transforms it into a model for the origin of translation in a distant RNA world without a supporting evolutionary tree to place this trait in the progenote (the most recent common ancestor of modern organisms). Moreover, it is quite likely that the trait is adaptive and therefore could have arisen in the modern world, as Maizels and Weiner have acknowledged elsewhere⁵.

Maizels and Weiner's disagreement with our focus on a 'breakthrough organism' is less than clear. In the context of the model, there must have been a first organism to contain a genetically encoded message which (given the homology of all modern ribosomes) lived before the progenote. This does not mean, nor have we written, that translation arose "all at once", or that translating and non-translating organisms did not coexist for an arbitrary time. But such a major metabolic innovation must have caused

extinctions in a metabolically and ecologically complex RNA world. This means that the breakthrough organism is (at best) the most ancient organism whose metabolism can be reconstructed simply by extrapolation from the biochemistry of modern organisms. Thus, models of the organism, together with prebiotic chemistry, must be the starting points for speculation about the RNA world.

A study of the chemical details of modern metabolism allows us to build these models in some detail^{2,6}. For example, deoxyribonucleotides are biosynthesized from ribonucleotides, suggesting that RNA evolved before DNA. However, three (or four) mechanistically distinct ribonucleotide reductases exist in the modern world⁷. In the model, this implies that the breakthrough organism used DNA to store genetic information⁸. This is yet another indication of the metabolic complexity of the last ribo-organism.

STEVEN A. BENNER

*Laboratory for Organic Chemistry,
Swiss Federal Institute of Technology,
Zurich, Switzerland*

ANDREW D. ELLINGTON

*Department of Molecular Biology,
Massachusetts General Hospital,
Boston, Massachusetts 02114, USA*

1. Maizels, N. & Weiner, A.M. *Nature* **330**, 616 (1987).
2. Benner, S.A. et al. *Cold Spring Harb. Symp. quant. Biol.* **52**, (1988).
3. White, H.B. III *J. molec. Evol.* **7**, 101-104 (1976).
4. Visser, C.M. & Kellogg, R.M. *J. molec. Evol.* **11**, 171 (1978).
5. Weiner, A.M. & Maizels, N. *Proc. natn. Acad. Sci.* **84**, 7383 (1987).
6. Benner, S.A. & Ellington, A.D. *Nature* **329**, 295 (1987).
7. Hogenkamp, H.P.C., Follmann, H. & Thauer, R.K. *FEBS Lett.* **219**, 197-201 (1987).
8. Benner, S.A. *Redesigning Life* 115 (Springer, Berlin, 1988).

Potential of ferroelectrics as radiation shields

SIR—In the commentary on the radiation hazards of space missions (*Nature* **330**, 709-710; 1987) Letaw et al. conclude that a 30-cm-thick aluminium shield reduces the dose by about 40% at solar minimum and by less than 25% at solar maximum. These estimates are obviously based on the well-known processes of loss of energy of charged particles in matter, which do not depend on the dynamics of solid state. But there is one solid-state effect with the potential of providing effective shielding without increasing the mass of the absorber.

In the case of ferroelectric targets of barium titanate and triglycine sulphate at their Curie temperatures, we have observed¹ a 2-3% increase in the energy loss of electrons of 620 and 942 keV. In their theoretical estimate of this anomalous energy loss at the Curie temperature, Ginsberg and Sobyannin obtained a value several times smaller but predicted a much larger loss in the case of the ferroelectric substance tris-sarcosine calcium chloride². Ayyub et al. observed about 12% increase in energy loss in one of their triglycine sulphate samples for 624-keV electrons³, and we observe an almost 100% increase

for 1.7-MeV electrons in barium titanate.

This anomalous increase in energy loss of charged particles in ferroelectrics at the Curie temperature is expected to become more pronounced at relativistic energies. Were this found to be so, ferroelectrics would have many advantages as a shielding material in space missions. There are many ferroelectrics with convenient Curie temperatures. Many of them have high hydrogen content and so can efficiently slow down fast neutrons, and some of them contain elements of high atomic number, which would be efficient gamma attenuators. Hence this process deserves to be investigated in greater detail.

M.T. LAGARE

*Department of Physics, Gulbarga
University,
Gulbarga, 585 106, India*

N. UMAKANTHA

*Department of Physics, Karnatak
University,
Dharwad, 580 003, India*

1. Lagare, M.T. & Umakantha, N. *Phys. Rev. Lett.* **49**, 1431-1433 (1982).
2. Ginsberg, V.L. & Sobyannin, A.A. *Soviet solid state Phys.* **25**, 1163-1164 (1983).
3. Ayyub, P. et al. *Phys. Rev. B* **32**, 2835-2840 (1985).

Size of the Sun in the seventeenth century

SIR—Our report¹ that the solar diameter increased during the second half of the seventeenth century was recently criticised² on the basis of two assumptions concerning the accuracy of the seventeenth-century observations made by Picard and his colleagues: that Picard had observed the polar diameter of Jupiter; and that the accuracy of the micrometer method developed by Auzout could not have enabled the detection of a 3 arc s variation of the diameters of the Sun and planets.

There is no evidence to substantiate such assertions, however. In his original manuscripts³, Picard did not state explicitly which diameter he was measuring, except in the case of the Sun and Saturn, where both the horizontal and vertical diameters were measured. In Picard's time, however, the usual procedure was to measure the horizontal diameter of a planet close to meridian transit, when it can best be seen and when refraction effects are minimal. The horizontal diameter would then be close to the equatorial diameter. Picard compiled elaborate tables of atmospheric refraction to deal with that problem, and measurements of equatorial diameter by Newton and Pound⁴ are in good agreement with Picard's results.

In relation to the accuracy of the filar micrometer method, Cassini measured the ellipticity of Jupiter in 1691. He found an oblateness of 2.7 arc s, which is close to the real value (3 arc s). It should be noted that Picard and Cassini used the same equipment. Their refractor was fitted with a filar micrometer. The micrometer screw was carefully calibrated by measuring the distance between nine hairs placed in the eyepiece. Thus, they could achieve an accuracy of 1 arc s on the apparent diameters of the Sun and planets⁵. This was discussed at length by Auzout in his treaty "Manière exacte" in 1667 (ref. 6). The document mentioned by O'Dell and Van Helden was a preliminary report to the Royal Society in December 1666 (ref. 7).

It seems likely, therefore, that it was the equatorial diameter of Jupiter that Picard measured. Comparing his observations only with polar diameters raises serious problems. The difference between Picard's observations and the actual equatorial diameters is only 1.3 ± 0.7 arc s. This discrepancy arises mainly from eight (out of eighteen) observations of Jupiter, made close to opposition. The difference can be explained by the "irradiation" effect: the contrast of a bright object, Jupiter, against a dark sky (near opposition) would result in an overestimation of the diameter. The irradiation effect was well known at the time of Picard, and because of this effect, most observers made their observations when Jupiter was in quadrature, 90° from the Sun-Earth

direction. If we consider the ten observations of Jupiter made by Picard that were not at opposition, the internal consistency of these observations is remarkable (see figure). The difference between Picard's diameters and the actual equatorial diameters is then reduced to -0.7 ± 0.4 arc s. This difference is probably due to the phase effects, whereby the equatorial diameter of Jupiter is slightly reduced by about 0.4 arc s near quadrature because the equatorial diameter, as seen from Earth, is not totally illuminated by the Sun. The curvature noted in Picard's observations near quadrature indicates this phase effect. Had his observations been corrected for the phase effect, the discrepancy between his estimate of Jupiter's diameter and the actual diameter would be negligible.

There is, therefore, no reason to add any correction to the night measurements made by Picard. The irradiation effect is smaller for the Sun than for Jupiter, as it is partially dependent upon the focal length of the refractor. The 6-ft quadrant used for the Sun was more favourable than the 14-ft refractor used for the planets. According to Lalande⁵, a systematic correction of 3–5 arc s should be applied to the seventeenth- and the eighteenth-century micrometer and transit observations if they are to be comparable with solar diameters derived from an eclipse. This correction reduces the solar diameter of Picard and La Hire to 32'6" or 32'4" during the deep Maunder minimum, and to 32'3" or 32'1" at the end of the Maunder minimum. The latter estimate is in agreement with the solar eclipse observed by Halley in 1715, also at the end of the Maunder minimum.

The correction of 3–5 arc s contains all possible sources of errors (optics, seeing, personal bias), and is quite compatible with our results. O'Dell and Van Helden have invoked a hypothetical systematic error on the transit timings. They suggest that a systematic error would remain even after smoothing out the individual fluctuation due to the quantization of time with a large number of observations. This is not possible. The accuracy of the pendulum used by Picard was better than two seconds per day, which results in an uncertainty of 0.02 arc s on the solar diameter. As discussed in our paper¹, the large error due to the quantization of time by the observer is divided by the square-root of the number of annual observations and leaves no systematic error. Our conclusions therefore remain unchallenged.

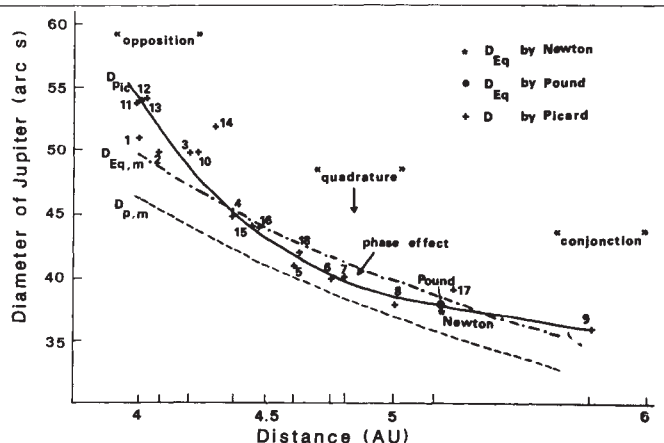
E. RIBES

Observatoire de Paris,
92195–Meudon, France

J. C. RIBES
R. BARTHALOT

Observatoire de Lyon,
69230–Saint Genis-Leval, France

The diameter of Jupiter observed by Picard (+) plotted against the distance between the planet and the Earth. Stars and circles refer to the equatorial diameter of Jupiter, observed by Newton and Pound, respectively, during the Maunder minimum. The modern values of the equatorial (dots) and polar (dashes) diameters of Jupiter are shown for comparison. Near quadrature, the observations made by Picard and his seventeenth-century colleagues are in good agreement with present-day estimates of the equatorial diameter. Close to opposition, the difference in Picard's observations is probably due to the irradiation effect.



O'DELL AND VAN HELDEN REPLY—Picard did not state which diameter of Jupiter he measured because it made no difference: first, he could barely verify that the planet is oblate but could not measure a difference between the polar and equatorial diameters; second, the differential atmospheric refraction in the case of Jupiter is negligible (0.04 arc s at an elevation of 30°), and Picard was aware of this since his refraction table did not vary much from the currently accepted one. Picard's telescopes had no clock-drives, and it was impossible to follow a planet (moving through the field at the rate of 15 arc s per second of time) smoothly enough to measure the horizontal diameter with any reasonable accuracy. The most natural way to measure a diameter was to hold the telescope firm and allow the planet to traverse the field, grazing the micrometer hairs. In the case of Jupiter, this meant measuring the polar diameter. Ribes *et al.* state that it was usual to measure a planet's horizontal diameter but offer no evidence for this statement.

Ribes *et al.* cite measurements by Newton and Pound and argue that these colleagues were also concerned with atmospheric refraction in this context. An examination of Book III, Proposition XIX of the third edition of Newton's *Principia* (1726, not 1694) reveals no mention of atmospheric refraction in this context.

More important, Newton did not measure Jupiter's diameter at all: he relied on measurements made by James Pound. Newton discussed these measurements in Book III, Phenomenon I, and it is very clear from this discussion that he was well aware of the problems posed by irradiation and micrometer errors. He based his conclusion that at mean distance Jupiter's apparent diameter is 37 arc s on Pound's timing of transits of Jupiter's satellites across the planet's disk. This method is insensitive to irradiation effects. Not quite as accurate were the micrometer measurements made by Pound with a 123-

foot telescope (whose aperture ratio approached f200), by means of which apparent diameter of 39 arc s was measured. With shorter telescopes, such as those used by Picard, Newton was of the opinion that Jupiter's apparent diameter would be measured to be 41 arc s. Newton's judgment is echoed by Lalande: seventeenth-century micrometer measurements were too large by three to five arc seconds, and this is precisely our point.

In our paper² we relied on Auzout's report to the French Academy, which was written at about the same time as his *Manière Exacte*. It is not a preliminary report but a slightly earlier report.

Ribes *et al.* present data to try to establish a variable irradiation correction depending on illumination of the sky, so that the correction would be smaller at quadrature than at opposition. If this effect were important, the correction should be even smaller near conjunction, yet Picard's measurement no. 9 on the figure, made near conjunction, indicates a larger correction than the measurements made at quadrature. We conclude that this condition of the observations is not important.

Finally, it appears that Ribes *et al.* simply do not understand the 'personal equation' in astronomical measurements. It is a non-statistical effect that varies from observer to observer and arises from systematic early or late timing of an event.

C. R. O'DELL
A. VAN HELDEN

Scientia, Rice University,
PO Box 1892,
Houston, Texas, USA

1. Ribes, E., Ribes, J. C. & Bartholot, R. *Nature* **326**, 52–55 (1987).
2. O'Dell, C. R. & Van Helden, A. *Nature* **330**, 629 (1987).
3. Picard, J. manuscripts D1, 14–16. *Archives de l'Observatoire de Paris* (1666–1682).
4. Newton, I. *Philosophiæ Naturalis Principia Mathematica*, LIII, Phenom. Prop. XIX, 3rd Edn (1694).
5. LaLande, J. *Astronomie*, Ed. Desainp. Paris, 2nd edn (1771).
6. Auzout, A. *Mémoires de l'Académie des Sciences (Paris)*, Tome VII, 118–130 (1729).
7. Auzout, A. *Phil. Trans. R. Soc.* **21**, 373 (1667).

Aspects of the major element composition of Halley's dust

Elmar K. Jessberger, Alexander Christoforidis & Jochen Kissel

Max-Planck-Institut für Kernphysik, PO Box 103980, 6900 Heidelberg, FRG

The chemical composition of the solid grains from comet Halley can be inferred from impact-ionization time-of-flight mass spectrometry. Halley's dust in the vicinity of the VEGA-1 spacecraft is a mixture of a refractory organic component and unequilibrated silicates, but detailed chemical information on individual particles is difficult to extract because of the complexity of the impact-ionization process.

THE determination of the chemical and possibly isotopic composition of cometary solid material was one of the main objectives of the three missions to comet Halley, VEGA-1, VEGA-2, and Giotto. The experiments to this end were the impact-ionization time-of-flight mass spectrometers, PUMA-1, PUMA-2 and PIA respectively¹. Immediately after completion of the missions, first results have been published²⁻³. Subsequently, reports on the average chemical composition⁴⁻⁸, on the compositional variability⁹⁻¹², on the isotopic composition of some elements^{7,13}, and on the organic component¹⁴ were given.

Evaluation

Here we describe further attempts to extract chemical information on the solid dust particles from mass spectra. We restrict ourselves to spectra from the experiment PUMA-1, as it functioned the best of all three experiments (compare ref. 6), and to the so-called mode-0 spectra that were transmitted from the space-probe to Earth in a transient-recorder-like mode and which are the least compressed spectra (for details of the other modes see ref. 1). The present study encompasses all 79 spectra of this kind with discernible mass peaks.

The evaluation of the measured spectra for chemical purposes proceeds by comparing them with synthetic spectra, calculated on the basis of normal isotopic composition, singly charged mono-atomic ions. For each element, the number of ions are calculated that can be accommodated by the measured mass peaks without conflict with normal isotopy (examples of this procedure are shown in refs (4) and (7)). Often the calculated spectrum resembles the measured one, but deviations exist, mostly notably in the case of carbon, where the measured 12/13 intensity ratios vary between 4 and 4,000 (refs 7, 13), and of magnesium, with 24/25 intensity ratios in the range 3-100 (ref. 7). Excesses of the heavy minor isotopes can be accounted for by hydrogenation, and excesses such as that on mass peak 24 are probably due to C₂⁺. The important but difficult question as to whether it is possible to extract 'real' isotopic compositions of some elements is discussed elsewhere^{7,13,15,16}.

We assume normal isotopy of the elements for the chemical analysis. As the presence of some molecular ions is much more probable than dramatic isotopic aberrations from normal, we would argue that the ion numbers resulting from the adopted procedure are actually upper limits. The next step is to convert the ion numbers to atom numbers. Because of the lack of any realistic calibration, we have used the relative ion yields derived for this type of instrument from secondary ion mass spectrometry (SIMS) yields and from calibration measurements made with particle speeds <64 km s⁻¹ (refs 17, 18). These are presumed to be accurate within a factor of two.

One of the main problems of the PUMA and PIA instruments, indeed, is their calibration, because there is no machine capable of accelerating any grains to 70-80 km s⁻¹, let alone insulating ones loaded with volatiles. Theoretical calculations on the impact-ionization yields resulted in conflicting predictions, and

Table 1 Average atomic abundances of the elements in Halley's dust and carbonaceous (CI) chondrites¹⁹

Element	Halley's dust	CI	Halley/CI	<i>P</i> %	
H†	2,025±385	492	4.1	0.5	51
C	814±165	70.5	11.6	5.9	84
N	42±14	5.6	7.5	ND	71
O	890±110	712	1.3	0.8	81
Na	10±6	5.3	1.9	ND	81
Mg	≡100	≡100	1.0	1.0	91
Al	6.8±1.7	7.9	0.9	ND	64
Si	185±19	93.0	2.0	1.6	91
S	72±23	47.9	1.5	1.2	73
K	0.2±0.1	0.35	0.5	ND	39
Ca	6.3±1.9	5.68	1.2	0.9	68
Ti	0.4±0.2	0.22	1.9	ND	15
Cr	0.9±0.2	1.25	0.7	ND	34
Mn	0.5±0.2	0.88	0.6	ND	27
Fe	52±9	83.7	0.6	0.5	73
Co	0.3±0.2	0.21	1.2	ND	6
Ni	4.1±2.1	4.59	0.9	ND	39

The data on cometary dust derived in this work are based on a total of 79 high quality mass-spectra from PUMA-1. For the derivation and the meaning of the quoted uncertainties see text. *P*, Percentage of mode-0 spectra in which element has been identified. ND, not determined.

* Calculated from an analysis of highly compressed PUMA-1 spectra⁶.

† From 'short' spectra only.

also do not agree with the experimental PUMA and PIA data. For example, the ratio ⁴⁰Ca²⁺/⁴⁰Ca⁺ was calculated as ~1 (J. Inagamon, personal communication), but the PUMA spectra¹⁶ indicate that ⁴⁰Ca²⁺ is virtually absent. In this peculiar situation, three expedients have been followed: (1) not to convert ion numbers to atom numbers or atomic ratios^{5-7,12}; (2) to assume that the bulk chemical composition of Halley's dust is the same as that of type 1 carbonaceous chondrites (ref. 11 and this report); (3) to use the ionization yields given by refs 17 and 18, remembering that the yields are thought to be accurate within a factor of two (refs 4 and 7 and this report). No corrections for 'matrix' effects and so forth are possible, simply because there is not enough input from calibration experiments. Expedient (1) is most suitable for discussion of isotopic ratios and temporal/spatial relative compositional variations, but only (2) and (3) allow the derivation of chemical information. Confirmation of the validity of the used ion yields would follow if they resulted in a complete interpretation of all data without any contradiction and within the framework of the chemical evolution of the early solar system derived from the study of meteorites and interplanetary dust particles.

Average composition

The sum of the atom numbers derived from all 79 spectra represents the average chemical composition of Halley's dust

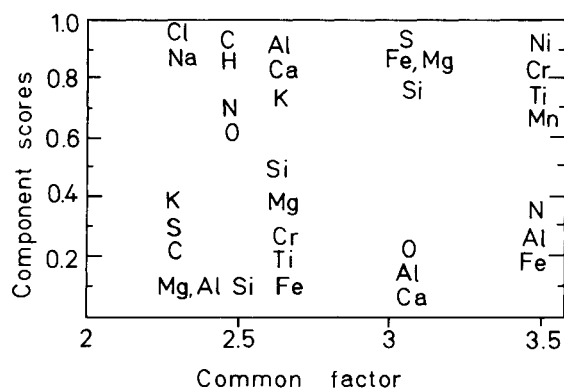


Fig. 1 Score chart of a common factor analysis of the elements in 79 Halley grains. Component scores below 0.1 are omitted. The higher the score of elements at a common factor, the more closely related they are.

grains. High intensity spectra ('large' grains) dominate this sum so that typically 60% of the total intensity of an element is contributed by 10% of the grains. Because of the differing characteristics of 'long' and 'short' spectra^{6,7,16}, average element ratios were calculated separately for the two types of spectra. [In the PUMA instruments two voltage settings alternate every 30 s to reflect ions with initial energies up to 50 eV (long spectra) and up to 150 eV (short spectra), respectively. Ions with larger initial energies hit the wall and are lost for analysis. The average Mg-normalized ion-abundances in both types of spectra are very similar for the rock-forming elements, while the Mg-normalized CHON-element ions are more abundant in short than in long spectra¹⁸.] For normalization we have chosen Mg instead of the commonly used Si: the isotopic control used to identify the elemental ion abundance is sharper for Mg (24/25/26 = 7.3:0.9:1) than for Si (28/29/30 = 29.7:1.5:1) because the minor isotopes of Mg are less likely to be affected by background than those of Si.

The second column of Table 1 gives the abundances of the elements in Halley's dust. This is a revised version of our earlier data^{4,7}, which were based on much fewer spectra. Two attempts were made to obtain a realistic measure of the statistical uncertainty of the element abundances shown; this is not a trivial task as there is no error figure available for the ion-number of an element in a spectrum (the problem of the methodical uncertainty, namely the yield problem, has been discussed already). The variations in the element ratios that occur, if each spectrum is omitted in sequence from summation of the atom numbers, were first calculated. This procedure resulted in errors in the range 3%–10%, which is considered unrealistically small. In the second attempt, long and short spectra were each divided into two random groups for summation and Mg-normalization. The geometric means from the four resulting groups are the quoted elemental abundances (Table 1), and the uncertainties are the standard deviations from the means.

Also in Table 1 are shown the element abundances in the least-altered early solar-system material at present available for laboratory analysis, the carbonaceous chondrites of type 1 (CI chondrites) (ref. 19). The average composition of both materials is the same within a factor of two, which coincides with the accuracy of the ion yields, but with the notable exception of the very light and volatile elements. The distinction of the H, C, N and O elements (oxygen is clearly related to CHN, see below and ref. 6) from 'silicate-elements' has been noted before^{2-4,7-10,12-14}. It gave rise to the identification of two end-member particle types, CHON-particles and silicates. The apparent enrichment of the elements H, C, N and O (Table 1) demonstrates that Halley's dust is richer in volatiles and more solar-like, and hence probably less altered and more primitive than early Solar System material available as CI chondrites⁴.

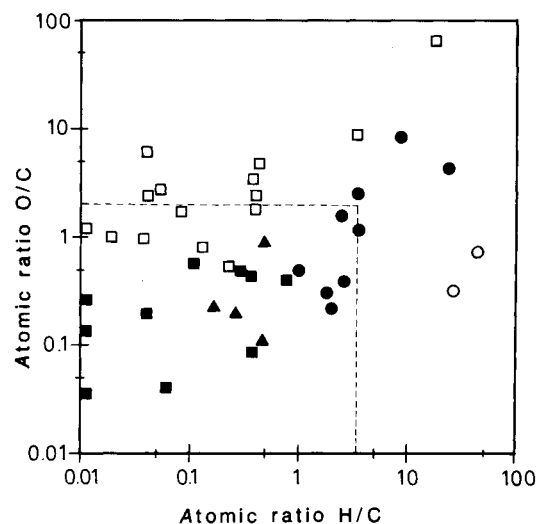


Fig. 2 O/C versus H/C (atom ratios) from 40 'short' high fidelity PUMA-1 spectra. Symbols represent the five classes identified by cluster analysis (see text): A, open circles; B, closed circles; C, open squares; D, closed squares; E, triangles. The field underneath the dashed line represents possible stoichiometric compositions of H, C and O. Higher O/C ratios may be due to oxygen from silicates, excessive H may come from the Ag target.

In the fifth column of Table 1 are listed the results of an analysis by Langevin *et al.*⁶ of 2,204 spectra transmitted in higher compressed modes from PUMA-1, after conversion using the same ion yields as in the present study. The isotopic-control procedure used in our present study to signify multi-isotope elements cannot be applied rigorously in the analysis of highly compressed mode spectra, and minor elements are not detectable. With the exception of hydrogen, for which discrepancy the data acquisition mode may be responsible, the agreement is excellent, given the different approaches and data sets. The similarities of the relative element abundances reflect the similarities of the respective ion ratios, thus lending credibility to the evaluation procedures.

The abundances of the rock-forming elements (Table 1) in Halley coincide with those in CI chondrites almost within the quoted uncertainties, irrespective of their cosmochemical group¹⁹, with the exceptions of Si and Fe. The apparent bulk Si enrichment and Fe deficiency, as shown by the Fe/Si atomic ratio of 0.3, is very uncommon in known primitive solar system materials²⁰. In the sun and bulk chondrites this ratio is close to one. It varies very little, although systematically, between different chondrite classes from 1.06 to 0.73 (ref. 20). Only in differentiated achondrites is the range larger, varying from 1.4 to 0.22. If the apparent ratio of 0.3 for Halley's bulk dust is true, the dust would appear to have an evolved²¹ silicate component. We cannot exclude this possibility at present and it may even constitute a major new and unexpected finding. Fe/Si-ratios of ~0.5 have been found for bulk interplanetary dust particles by electron²²- and proton²³-induced X-ray analysis. But as the deviations of Si/Mg and Fe/Mg from the respective CI-chondrite ratios are well within the uncertainty of the ion yields used^{17,18} it is probably safer to attribute the deviations to them, rather than to propose a revolutionary model for the origin of comets. But we may have missed a very important finding!

Groupings

To find out which elements belong together, we have performed a common factor analysis²⁴ with the absolute numbers of atoms on each element in the grains. The resulting factor pattern is shown in Fig. 1. The elements arrange themselves into five groups: (1) Ni, Cr, Ti, Mn; (2) S, Fe, Mg, Si; (3) Al, Ca; (4) C, H, N, O; (5) Cl, Na. To comment on the last group first,

Table 2 Geometric mean elemental compositions (atom%) of the four groups of Halley's dust particles identified by cluster analyses

Element	Empirical dust group			
	A	B	C	D
C	54.4 ± 8.7	9.7 ± 4.0	1.8 ± 0.9	7.9 ± 4.6
N	3.1 ± 1.1	2.3 ± 1.0	<1	2.1 ± 0.7
O	11.6 ± 5.8	53.9 ± 7.1	1.8 ± 0.9	4.1 ± 2.3
Mg	4.1 ± 2.0	6.7 ± 3.4	41.5 ± 6.9	8.6 ± 5.3
Al	<1	<1	2.4 ± 0.8	1.5 ± 0.4
Si	3.1 ± 1.2	6.2 ± 2.6	28.6 ± 6.4	5.4 ± 2.2
S	2.4 ± 0.9	1.9 ± 0.7	<1	3.4 ± 1.7
Ca	<1	<1	3.3 ± 1.5	<1
Fe	2.6 ± 1.3	2.2 ± 1.0	5.1 ± 2.7	32.7 ± 11.1
Number of grains	37	18	10	11

Uncertainties reflect variation of element concentration within groups.

chlorine could be a surface contaminant. (F. Bühler, private communication), as indicated by SIMS analysis of the PUMA-1 Ag target. For this reason Cl has not been included in Table 1, where it would appear as >200-fold Cl chondrites. Whether Na, as the element accompanying Cl, is affected by contamination is not known. The CHON elements readily group together and are distinct from the rock-forming elements of groups 1–3. It is not clear why oxygen is only weakly related to the later elements, which are probably present mainly in the form of oxides, and is related instead to the carbonaceous component^{4,14}. The weak relation of oxygen with Mg and Si is not predicted from calibration measurements at lower velocities, and is an unexpected result of the space experiment. It is noteworthy that the refractories Ca and Al are grouped together, as are the major elements S, Fe, Mg, and Si: the latter grouping suggests that Fe is mostly present in the silicates, and not as Fe-metal particles adhering to the grains for example²⁵.

To separate groups of grains of similar chemical composition we have performed cluster analyses²⁴. Elemental variations are investigated by giving the particles equal weight using elemental atom percentages. The procedure of dividing the number of atoms of an element by the sum of all atoms in that grain is common practice in 'terrestrial chemistry', where the quality of data does not vary too much from entry to entry, or where the variation can easily be taken into account. Even with the highest quality data from Halley to form the basis of the present study, this simple transformation could profoundly affect the result. For example, a spectrum with low peak intensities does not allow an isotope control procedure as rigorous for element identification as a high-intensity spectrum. But this low-intensity spectrum could still be extremely important if it stems from a monomineralic grain. In the absence of clear-cut valuation criteria, we give the particles equal weight.

The result of the cluster analyses is given in Table 2. Four main groups are distinguished:

- A: C, O rich; $C/O = 4.7 \pm 2.5$; $(C+O)/(Mg+Si+Fe) = 6.7 \pm 2.1$
 B: C, O rich; $C/O = 0.2 \pm 0.1$, $(C+O)/(Mg+Si+Fe) = 4.2 \pm 1.3$
 C: Mg, Si rich; $(C+O)/(Mg+Si+Fe) = 0.05 \pm 0.02$
 D: Mg, Fe rich; $(C+O)/(Mg+Si+Fe) = 0.26 \pm 0.13$

Given ratios are calculated from the geometric means in Table 2. CHON particles dominate groups A and B, which are distinguished by their C/O ratios. Groups C and D comprise the silicate grains. The Si/Mg ratios of the mean atom concentrations in group C and D are very similar (~0.65), whereas the Fe/Mg ratios are different (C: 0.12; D: 3.8).

CHON particles

Kissel and Krueger¹⁴ have reported on the composition of the hydrocarbons in Halley's dust inferred from a composite spec-

trum of what remained after the singly charged monatomic ions had been removed. It should be stressed that the organic component must be refractory because the survival of ices can be thermodynamically excluded at the Halley-VEGA-1 distance²⁶.

Here we attempt to classify individual grains on the basis of the elemental concentrations of H, C, N and O. We have to restrict ourselves to 'short' spectra, because in 'long' spectra H is evidently deficient on experimental grounds^{6,16}. As the quantitative determination of H in mass spectrometry is generally difficult, and as H may be released from the Ag target on impact (F. R. Krueger, private communication), the H data should be viewed as substantiated estimates. Table 3 gives the mean concentrations of H, C, N and O in five CHON groups, which are identified by cluster analysis. The smallest group, A, is dominated by hydrogen. It apparently contains more H than can be accounted for by organic stoichiometry. In group B, hydrogen is accompanied by almost equal amounts of carbon and oxygen. Groups C, D and E are much lower in hydrogen. Groups D and E are both carbon-rich and differ mostly in the C/O ratio. Group C, the largest group, is oxygen-rich.

The CHON groups are also shown in an O/C versus H/C diagram (Fig. 2). H/C ratios of the hydrogen-rich groups A and B are well above the approximate lower limit set by most hydrocarbons. Groups D and E, which are almost indistinguishable in this plot, populate the C-rich lower left corner. Group C may indicate in which grains oxygen is associated with CHON or with the silicates: most of the high O/C and low H/C ratios are accompanied by a high (Si+Mg) abundances in which case some oxygen is likely to be associated with the silicates.

The range of element ratios observed for Halley's dust is greater than observed elsewhere. High relative carbon abundances can easily be accounted for by the presence of amorphous carbon or, more exotically, interstellar diamonds²⁷. But the large O/C and H/C values appear to be excessive. The maximum H/C observed in interstellar molecules is 5 and the maximum O/C is 2 (ref. 28). The presence of polycyclic aromatic hydrocarbons²⁹ and polyoxymethylenes³⁰ is inferred in interplanetary dust particles and Halley's dust, respectively; both would populate rather small areas in Fig. 2. The elemental variations among the CHON-rich grains have not yet been interpreted in detail, but they will inevitably set limits for the composition of Halley's refractory organic component.

Silicates

Some aspects concerning the major rock-forming elements will now be discussed. We have attributed the deviations of the Si/Mg and Fe/Mg ratios from the respective CI-chondrite ratios to inadequately known ion yields. As a consequence we correct the ion yields so that the average Mg-normalized abundances of the non-CHON elements are the same as in CI chondrites. The same procedure has also been adopted by Brownlee *et al.*¹¹. The changes are within the uncertainty of the given ion yields^{17,18}, but do not follow the systematics that could be expected from following Krueger and Kissel¹⁸. We are aware that this procedure does not exclude circular reasoning.

Table 3 Geometric mean composition (atom%) of five groups of dust particles identified by cluster analysis of the data from 40 'short' spectra of PUMA-1

Element	Empirical CHON group				
	A	B	C	D	E
H	94.1 ± 0.3	50.0 ± 4.3	5.0 ± 2.7	2.6 ± 1.4	5.3 ± 2.8
C	2.7 ± 0.4	12.4 ± 4.7	11.2 ± 4.6	64.8 ± 3.8	26.7 ± 7.7
N	<1	1.8 ± 0.5	3.1 ± 1.6	5.0 ± 1.9	3.2 ± 0.9
O	1.3 ± 0.2	15.7 ± 4.5	50.4 ± 4.8	6.3 ± 2.0	16.4 ± 5.3
Number of grains	2	9	15	10	4

Uncertainties reflect variation of element concentration within groups.

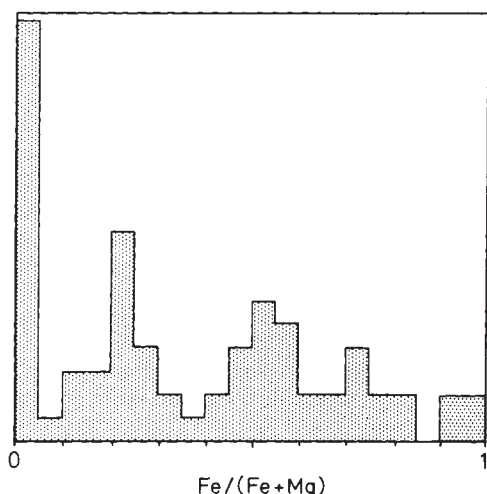


Fig. 3 Histogram of the $\text{Fe}/(\text{Fe}+\text{Mg})$ atom ratios in 74 grains of Halley's dust.

It had been inferred that a number of Halley's dust grains are mostly FeS (refs 6, 12), a common sulphur-bearing compound in meteorites. Forty-six out of the 79 grains considered here contain both Fe and S. Three of the grains have sulphur contents larger than 10 atom% and Fe/S ratios in the range 0.5–2. Thus only a few FeS-dominated grains can be discerned.

Brownlee *et al.*¹¹ compared the submicron-distribution of $\text{Fe}/(\text{Fe}+\text{Mg})$ in the matrix of carbonaceous chondrites with the distribution they obtained for Halley's dust. They found that "the distribution of Mg and Fe in Halley is unlike that seen in meteorites" and concluded that "Halley-like material has not been preserved in chondrites". The distribution of $\text{Fe}/(\text{Fe}+\text{Mg})$ in chondrite matrix was found to be rather narrow, typically 0.3–0.7, and strongly centred around 0.5, whereas in Halley it ranged from 0 to 1 without any marked grouping. The data from the present study are shown in Fig. 3 and were obtained with the aforementioned isotopic control which identified Mg and Fe on the basis of their isotopic composition rather than just the main isotope peak height, which could be subject to molecular interferences⁷. These data encompass more spectra than those in ref. 11. As in ref. 11, the ratios cover the whole range 0 to 1, but in contrast to those results¹¹, some regimes are well separated to give the two extremes with zero Fe (18 grains) and low Mg (4 grains), and two intermediate groups with $\text{Fe}/(\text{Fe}+\text{Mg}) \sim 0.2$ and ~ 0.5 , respectively. In a recent study³¹ such a plot is presented that was obtained from point count analyses of two interplanetary dust particles (IDPs) and the matrix of the carbonaceous chondrite Orgueil. The $\text{Fe}/(\text{Fe}+\text{Mg})$ distribution in a layer silicate IDP and in Orgueil matrix (see also ref. 11) are very similar and peak strongly at around $\text{Fe}/(\text{Fe}+\text{Mg}) = 0.5$, whereas an anhydrous IDP has a strong maximum at very low ratios and is smeared out up to $\text{Fe}/(\text{Fe}+\text{Mg}) = 1$. The wide range and the strong maximum at $\text{Fe}/(\text{Fe}+\text{Mg}) < 0.1$ in Halley's dust (Fig. 3) resemble more closely the distribution in the anhydrous than in the layer silicate IDP. The marked difference in the $\text{Fe}/(\text{Fe}+\text{Mg})$ distribution between cometary dust and carbonaceous matrix which is independent of ion yield ambiguities is also evident. One may speculate from the present data that layer silicates occur among Halley's dust in addition to anhydrous silicates; after all, there is a group at $\text{Fe}/(\text{Fe}+\text{Mg}) \approx 0.5$ (Fig. 3). It should also be noted that the composition estimated for Halley's dust from infrared data³² is 56% olivine, 36% pyroxene and 8% layer lattice silicates, which does not apparently contradict such a speculation. A test for the presence in Halley's dust of layered silicates, which are always hydrated, would be the identification of OH-rich and C-free grains. This test, however, is difficult because of the complex mixture of silicates and CHON.

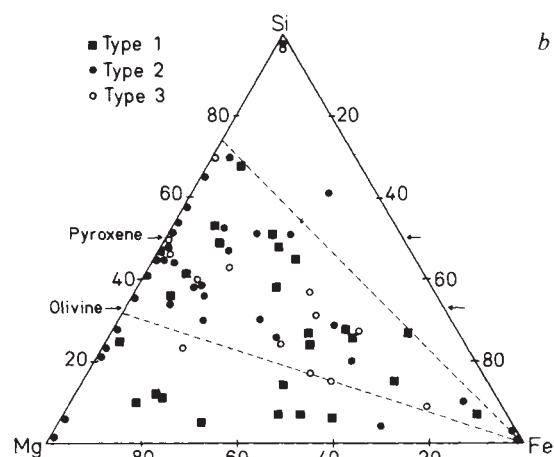
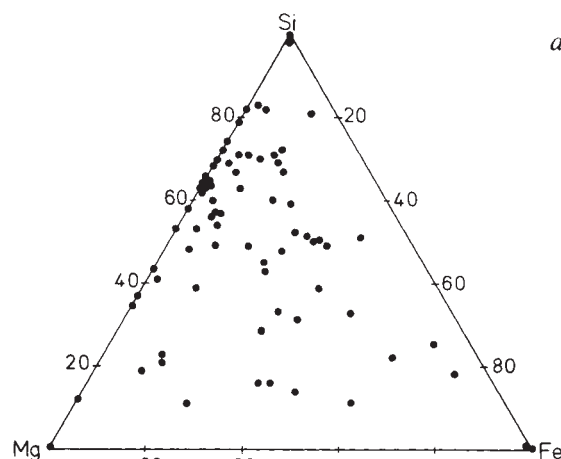


Fig. 4 Mg/Si/Fe in Halley's dust grains. *a*, Data evaluated with the ion yields given by ref. 17. *b*, Ion yields adjusted so that on average Mg/Si/Fe is chondritic (see text). The 'types' are a measure of the quality of the spectra, type 1 (filled squares) being highest, followed by type 2 (filled circles) and type 3 (open circles). Arrows indicate the loci of olivines and pyroxenes. The dashed lines span the triangle that encompasses more than two thirds of the grains.

Major element data are commonly displayed as ternary graphs. Figure 4 shows Halley's dust data from the present study in the system Mg-Si-Fe (atom%). To demonstrate the effect of the adjustment of the ion yields of rock-forming elements described above such that the average composition of Halley's dust is chondritic, the non-adjusted (Fig. 4*a*) and the adjusted (Fig. 4*b*) data are shown. As a result of the adjustment, the data points are moved towards Mg and Fe and are further spread out. Note that the number of grains closer to the pyroxene- and olivine-trend lines increases due to that effect. In Fig. 4*b* the data are graded according to the subjective quality criteria of the overall appearance of the spectrum and the clarity of the assignments of Mg, Si and Fe: type 1 being excellent and type 3 reasonable spectra. The absence of a relation of composition to the quality type in Fig. 4*b* supports the contention of unbiased data evaluation.

The scatter of the compositions is considerable (Fig. 4*b*), and much larger than in the submicron volumes of meteorites and interplanetary dust particles analysed by Wheelock and Brownlee³³. On the other hand, the data points in Fig. 4*b* do not appear to be randomly distributed: the triangle *ABC* (*A*: Si = 100; *B*: Mg = 75, Fe = 0; *C*: Fe = 100) is practically devoid of data, implying the lack of both Fe-rich pyroxenes and Fe-rich olivines. On the other hand, the triangle *BCD* (*D*: Mg = 35, Fe = 0) encompasses

more than two-thirds of all grains, 54 out of 75. The group of grains close to Fe = 0, Mg = 50, Si = 50 indicates the survival of primary Mg-rich silicates, which in turn may be taken as evidence for a 'low-temperature' history of the grains, and hence for the primitive nature of Halley's solid dust.

It has been argued from meteor spectroscopy³⁴ that some of the particles may consist of refractory minerals which resemble Ca, Al-rich inclusions found in certain carbonaceous chondrites and are believed to retain records from the earliest solar system. In the data set analysed in this study, only a few grains with a possibly high Ca + Al content were encountered, but in terms

of spectral quality they belong to type 3 and consequently should not be ascribed too great a significance.

Presently the spectra transmitted in the higher compressed modes from the PUMA 1 and 2 and PIA experiments are being evaluated in greater detail; we intend to characterize these spectra by comparing them with the mode-0 PUMA-1 spectra of the present study and also to apply cluster analytical techniques.

We thank Eva Gensel for artwork. Criticism and advice from J. Bradley, E. Grün, E. Anders, D. E. Brownlee and F. R. Krueger are gratefully acknowledged.

Received 28 October 1987; accepted 3 February 1988.

- Kissel, J. *Eur. Spa. Ag. Spec. Pub.* **1077** 67-68 (1986).
- Kissel, J. *et al. Nature* **321**, 280-282 (1986).
- Kissel, J. *et al. Nature* **321**, 336-338 (1986).
- Jessberger, E. K., Kissel, J., Fechtig, H. & Krueger, F. R. *Eur. Spa. Ag. Spec. Pub.* **249**, 27-30 (1986).
- Wäsch, R. *Eur. Spa. Ag. Spec. Pub.* **250**, 265-267 (1986).
- Langevin, Y., Kissel, J., Bertaux, J.-L. & Chassefiere E. *Lunar Planet. Sci. Conf.* **18**, 533-534 (1987).
- Jessberger, E. K. & Kissel, J. *Lunar planet. Sci. Conf.* **18**, 466-467 (1987).
- Jessberger, E. K., Kissel, J., Fechtig, H. & Krueger, F. R. in *Physical Processes in Comets, Stars and Active Galaxies* (ed. Hillebrandt, W., Meyer-Hofmeister, E. & Thomas, H. C.) 26-33 (Springer, Heidelberg, 1987).
- Sagdeev, R. Z. *et al. Eur. Spa. Ag. Spec. Pub.* **250**, 349-352 (1986).
- Clark, B., Mason, L. W. & Kissel, J. *Eur. Spa. Ag. Spec. Pub.* **250**, 353-358 (1986).
- Brownlee, D. E., Wheelock, M. M., Temple, S., Bradley, J. P. & Kissel, J. *Lunar planet. Sci. Conf.* **18**, 133-134 (1987).
- Mukhin, L. M. *et al. Lunar planet. Sci. Conf.* **18**, 674-675 (1987).
- Solc, M., Vanysek, V. & Kissel, J. *Eur. Spa. Ag. Spec. Pub.* **250**, 373-376 (1986).
- Kissel, J. & Krueger, F. R. *Nature* **326**, 755-760 (1987).
- Solc, M., Jessberger, E. K., Hsiung, P. & Kissel, J. *Proc. 10th Regional Meeting of the IAU, Prague* (in the press).
- Jessberger, E. K., Rahe, J. & Kissel, J. in *Origin and Evolution of Planetary and Satellite Atmospheres* (ed. Matthews, M. S.) (University of Arizona Press, in the press).
- Kissel, J. & Krueger, F. R. *Appl. Phys. A*, 69-85 (1987).
- Krueger, F. R. & Kissel, J. *Naturwissenschaften* **74**, 312-316 (1987).
- Anders, E. & Ebihara, M. *Geochim. cosmochim. Acta* **46**, 2363-2380 (1982).
- Mason, B. in *Handbook of Elemental Abundances in Meteorites* (ed. Mason, B.) 209-213 (Gordon & Breach, New York, 1971).
- Anders, E. *Space Sci. Rev.* **3**, 583-714 (1964).
- Fraundorf, P. *Geochim. cosmochim. Acta* **45**, 915-944 (1981).
- Wallenwein, R., Blank, H., Jessberger, E. K. & Traxel, K. *Analytica chim. Acta* **195**, 317-322 (1987).
- SAS Institute Inc. *SAS User's Guide: Statistics*, Version 5, ed. (Cary, NC: SAS Institute Inc., 1985) 956 pp.
- Anders, E. *Eur. Spa. Ag. Spec. Pub.* **249**, 31-40 (1986).
- Hanner, M. *Icarus* **47**, 342-350 (1981).
- Lewis, R. S., Ming, T., Wacker, J. F., Anders, E. & Steel, E. *Nature* **326**, 160-162 (1987).
- Irvine, W. M., Schloerb, F. P., Hjalmarsen, A. & Herbst, E. in *Protostars and Planets II* (eds Black, D. C. & Matthews, M. S.) 579-620 (University of Arizona Press, Tucson, 1985).
- Allamandola, L. J., Sandford, S. A. & Wopenka, B. *Science* **237**, 56-59 (1987).
- Huebner, W. F. *Science* **237**, 628-630 (1987).
- Bradley, J. P. *Geochim. cosmochim. Acta* (submitted).
- McKeegan, K. D., Swan, P., Walker, R. M., Wopenka, B. & Zinner, E. *Lunar planet. Sci. Conf.* **18**, 627-628 (1987).
- Wheelock, M. M. & Brownlee, D. E. *Elemental Composition of Submicron Volumes of Meteoritic Material: Analogs to Cometary Dust Grains* (The Dept. of Astronomy, Univ. of Washington, Seattle, 1986) No. MW9/85.
- Wilkening, L. L. *Nature* **258**, 689-690 (1985).

Large-scale mantle convection and the aggregation and dispersal of supercontinents

Michael Gurnis

Seismological Laboratory, California Institute of Technology, Pasadena, California 91125, USA

The first time-dependent numerical simulations of continental aggregation and dispersal demonstrate a dynamic feedback between the motion of continental plates and mantle convection. Plate velocity is intrinsically episodic. Continental plates aggregate over cold downwellings and inhibit subduction and mantle cooling; the mantle overheats and fragments the continent under tension. Overall, the models are in agreement with the present geophysical state of the mantle and the geological record over the last 200 million years.

THE breakup of Pangaea which started 180 Myr ago is the dominant event of the late Phanerozoic. Before this breakup, Gondwana and the two northern continents (North America and Eurasia) had assembled together to form the Pangaeian supercontinent¹. Although the evidence is not clear cut, geologists are piecing together a picture of continents aggregating and dispersing a number of times over the last 3,000 Myr (ref. 2 and P. F. Hoffman and S. A. Bowring, unpublished manuscript). How does such aggregation and dispersal take place? It is almost certain that continental motions are tied to a large-scale mode of thermal convection. A physical model of continental aggregation and dispersal has been proposed by Anderson³ from continental reconstructions, the geoid and from hotspot locations. The supercontinent 200 Myr ago was approximately centred over the Atlantic-African geoid high, but North and South America, Australia, Antarctica and India moved away after breakup. Because this geoid high contains most of the world's hot spots³⁻⁵ and is characterized by low seismic velocities in the lower mantle, it is probably hotter than average^{6,7}. The continents (except for Africa) now appear to be moving towards

geoid lows, a low incidence of hotspots, and over lower mantle of high seismic velocity, factors which all indicate cold mantle. These observations suggest that supercontinents could affect the thermal state of the underlying mantle. The mantle underlying the continent becomes hotter than normal, expands and gives rise to geoid highs. This is followed by rifting in the supercontinent and the formation of hotspots; the supercontinent fragments and disperses³.

The feedback between continental plate and mantle convection is fundamental to this model: continental plates insulate the mantle, leading to high mantle temperatures; the perturbed mantle in turn causes continental motions. This two-way interaction has not been investigated in the few model studies of plates and convection. In general, modelling studies can be categorized as either imposing plate characteristics on the top of a convecting region and calculating the effects on convection^{8,9} or solving for convection in a region and calculating its effects on the motion of a plate^{10,11}.

The first dynamically self-consistent numerical models meant to simulate continental plates are presented here; they allow for

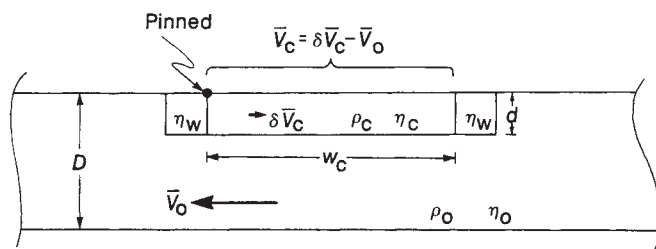


Fig. 1 Schematic of the continental plate within the interior of the convecting fluid. D is the depth of the convecting region. The continent is of viscosity η_c , density ρ_c , width w_c and depth d ; other parameters are defined in the text. For the calculations presented here, $\eta_c/\eta_0 = 10^3$, $\eta_w/\eta_0 = 10^{-1}$, and $d/D = 0.125$. The width was a free parameter. Because there were 32 equally sized finite elements in depth, the continent was 4 elements thick. The density difference, $\rho_0 - \rho_c$, was proportional to 20% of the density difference associated with the change in temperature across the box, that is $\rho_0 - \rho_c = 0.2 \rho_0 \alpha \Delta T$, where α is the coefficient of thermal expansion and ΔT the total vertical change in temperature.

a two-way interaction between internal convection and unsteady plate motion. The case of a single plate floating on top of the convecting region is studied, and the analysis is extended to two plates that can heal into a single plate, such as a supercontinent, and subsequently break up into two independent continents. The results for a single plate show that convection velocity is intrinsically episodic so that the continental plate experiences short bursts of rapid velocity. When a 'breakable' long plate is introduced, it will remain stationary, and the underlying fluid will heat up and break the plate into two fragments that rapidly move away from the hot area. The periodicity of this behaviour is ~ 100 – $1,000$ Myr, the variation in horizontal compressive and tensile stresses are ~ 10 – 10^2 MPa, peak plate velocities are ~ 5 cm yr $^{-1}$, and the subcontinental lithosphere temperature varies over a few hundred degrees Celsius. This insulation-induced heating mechanism, which gives rise to large-scale horizontal temperature variations, is in strong agreement with the geological record.

Model parameters

Thermal convection is calculated using the finite element method in a two-dimensional long box; the fluid is incompressible, has an infinite Prandtl number and a Rayleigh number of 10^5 . The code, which solves for the temperature equation explicitly and the Stokes equation with a penalty function formulation, has been extensively compared with other codes¹². The box has an aspect ratio (length/depth) of eight, and periodic boundary conditions; this means that fluid can flow out of one side of the box and into the other. For simplicity, the temperatures on both the base and top are held constant, a constant background viscosity is used and there is no internal heating in the fluid or plates. Without plates, this gives rise to Rayleigh-Bénard convection, the best understood mode of thermal convection treated by solid-Earth geophysicists¹³.

A continental plate is introduced by assigning high viscosity and an intrinsic chemical buoyancy to a group of elements along the top. The thermal properties are constant throughout the whole system. As illustrated in Fig. 1, the following variables are defined: the non-dimensional width and depth of the continental plate (w_c/D and d/D) and the density and viscosity of the continent (ρ_c/ρ_0 and η_c/η_0). The two sides of the continent are bounded by zones of low viscosity, η_w/η_0 . The top and bottom surfaces have free-slip boundary conditions, except for one pinned node that defines one corner of the continental plate. The average horizontal velocity through the box, excluding the continent, is $\bar{V}_0$, and is an output parameter of the calculation. If the continent had zero velocity, then the continental velocity with respect to the fluid would be $-\bar{V}_0$; this would be the case

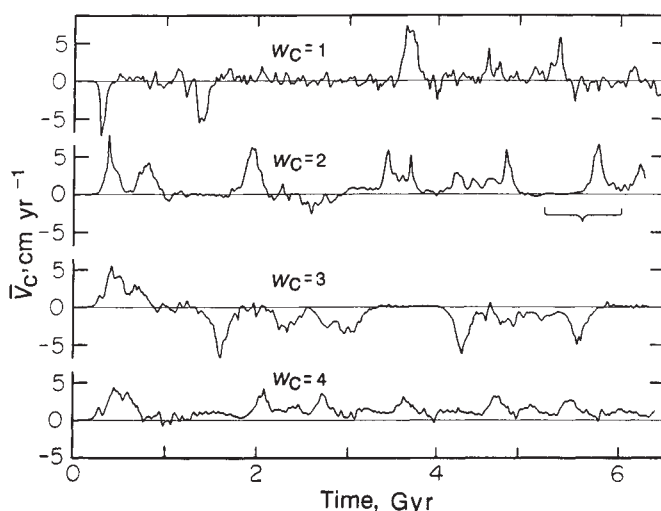


Fig. 2 Continental plate velocity $\bar{V}_c$, versus scaled model time, for four plate widths, w_c . The scaling parameters are given in Fig. 4. Each model was run for 14,000 'Courant' time steps.

if all of the nodes in the continent were pinned or if $\eta_c/\eta_0 \rightarrow \infty$. Because neither case is true, however, there are small velocity variations throughout the plate and consequently there is a residual velocity, $\delta \bar{V}_c$. The continental velocity is thus $\bar{V}_c = \delta \bar{V}_c - \bar{V}_0$. This may appear to be overly complex, but if there are several continents then only one can be pinned and the other(s) must move with respect to the pinned node, which is the second problem to be studied here. It turns out that for $\eta_c/\eta_0 = 10^3$, $\delta \bar{V}_c/\bar{V}_0 < 10^{-3}$. This formulation, with the continent pinned, allows a great reduction in the computation time. The stiffness matrix in the finite element formulation always remains the same and it need be formed and inverted only at the start of the time-dependent problem.

The case of two continents requires a few changes from the one-plate case. One plate is formed, as previously, with one node pinned, but the second plate must be advected. The average horizontal velocity is calculated in the second plate and the 'stencil' defining which elements have η_c and ρ_c is advected with this velocity. As the viscosities of the elements change during the model run, the stiffness matrix must be reformed and inverted; this reforming becomes more frequent when two plates are moving rapidly with respect to each other, and does not occur when the plates are joined together. Criteria must be defined for breaking one long plate into two short ones and vice versa. First, the areal average of the stress, σ_{xx} , is calculated from the velocity field. A plate is broken when σ_{xx} exceeds a critical yield stress, σ_{yld} , in which case each edge of the single plate moves slightly outwards and the centre high viscosity, η_c , is replaced by low viscosity, η_w . For simplicity, the low-viscosity zones are retained on the edges of the plate. When two plates are converging and the two low-viscosity zones overlap, the low viscosity, η_w , is replaced by η_c and the plate heals.

The intention of this model problem is not to simulate the entire plate tectonic-mantle convection system. Rather, the intention is to study one important feedback mechanism between plates and convection. Consequently, some aspects have been simplified from the true Earth, in particular the fact that in the oceanic areas there are no stiff plates. This is not a great drawback, however, because the oceanic areas still retain the fundamental property of having convection extending directly to the surface (oceanic lithosphere) and having the fluid that is cooled on the surface returned to the interior (subduction). The distinction between continents and oceans in this case is a 'conducting lid' on part of the fluid and free convection on the other part. Unique to this study is that the 'conducting lid' (continental

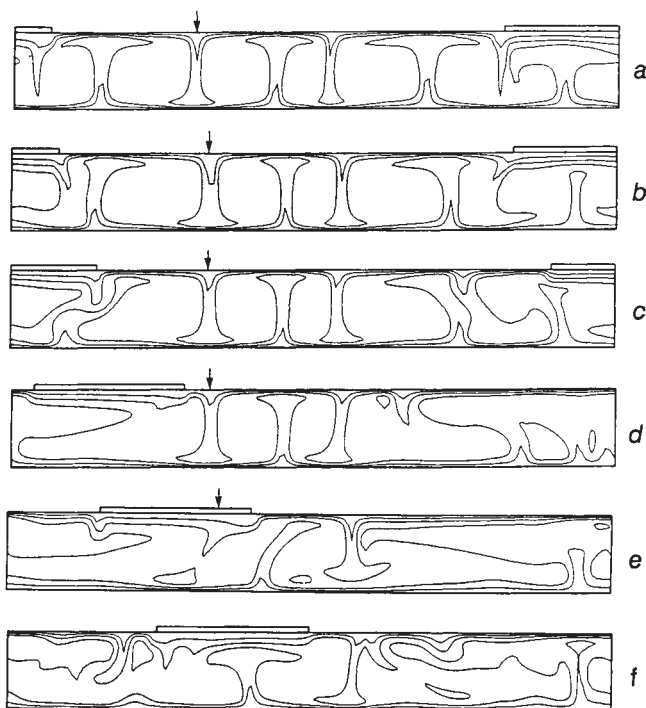


Fig. 3 Isotherms at four instants in a simulation involving a single plate of width, $w_c = 2$. The narrow box on the top of the fluid denotes the location of the plate; the height of the box has no physical significance as the continental plates are actually within the fluid. The vertical arrows denote a cold downwelling towards which the continental fragment moves. The Rayleigh number for the calculations was 10^5 . This figure was formed by computing a reference offset which was integrated forward with $\bar{V}_c$. This offset was then added to the raw temperature field and plate position.

plate) can go anywhere and take any velocity it chooses, allowing us to study some important feedback mechanisms that may cause continental drift. The next generation of these dynamic simulations will require both oceanic and continental plates, perhaps generated by a stress- and temperature-dependent rheology.

Single plate

Much of the physics experienced by breaking and healing plates will be contained in the motion of a single plate; for this reason single plates will be considered in detail here and the breaking-healing plates are presented as an extension to the present analysis. Through the rest of this article, the non-dimensional results are scaled to the Earth. This involves both scaling to a higher Rayleigh number as well as dimensionalizing using appropriate mantle values (see legend to Fig. 4); considering both aspects of the scalings, the final values (for example, the plate velocity, stress and basal temperature) are uncertain and are only accurate to within a factor of two. In Fig. 2 the velocity, $\bar{V}_c$, of single plates is presented for four widths, w_c , between 1 and 4. As seen in Fig. 2, plate velocity is not constant in any of the four cases; instead the velocity is distinctly episodic. For $w_c \leq 2$, periods of zero or low velocity are interrupted by short bursts of high velocity.

The cause of this velocity episodicity can be understood by carefully tracing the thermal and stress history of the fluid and plate through one cycle of the velocity pattern. Isotherms are shown in Fig. 3 and plate stress, velocity and temperature in Fig. 4 for $w_c = 2$. Starting with Fig. 3a, the continental plate is trapped between two cold downwellings, and directly over an upwelling. The plate is stationary and has zero average horizontal stress, σ_{xx} (Fig. 4a). The average temperature along the base of the plate increases linearly by about 1°C every 2.5 Myr for ~ 450 Myr (Fig. 4b), while the plate remains stationary. As

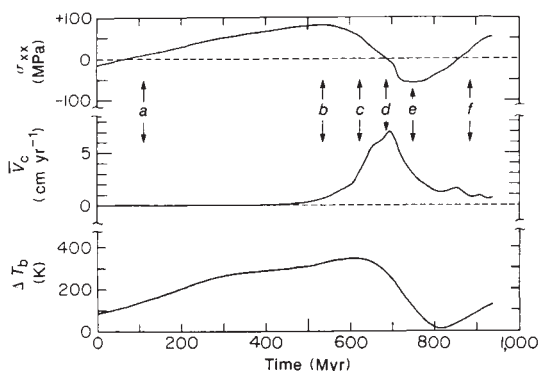


Fig. 4 Average horizontal stress through the plate, σ_{xx} , plate velocity, $\bar{V}_c$, and the change in temperature beneath the plate ΔT_b , for a single plate of width $w_c = 2$, as shown in Fig. 3. The letters labelling instants in time correspond to the isotherm plots in Fig. 3. The following parameters were used to scale the non-dimensional model results: a mantle depth of 3,000 km, an average mantle density of $4,000 \text{ kg m}^{-3}$, an average viscosity for the whole mantle of $5 \times 10^{21} \text{ Pa s}$, a total mantle temperature drop of 2,800 K, a coefficient of thermal expansion of $2 \times 10^{-5} \text{ K}^{-1}$, and a thermal diffusivity of $10^{-6} \text{ m}^2 \text{ s}^{-1}$. Parameters from ref. 23. These parameters give a Rayleigh number of 6×10^7 . Preliminary Rayleigh number scalings were determined by obtaining $\bar{V}_c$, σ_{xx} and T_b at $10^4 \leq \text{Ra} \leq 10^{5.5}$; $\bar{V}_c$ and σ_{xx} were written 10 per cent of the values obtained through boundary layer scalings²³; T_b has a complex functional form, but is approximately independent of Ra at $\text{Ra} > 10^5$. For the scalings, the boundary layer equations²³ were used for $\bar{V}_c$ and σ_{xx} and T_b is assumed independent of Ra. The non-dimensional values were then scaled using the above mantle parameters; the final scaled $\bar{V}_c$, σ_{xx} and T_b are only accurate to within a factor of two.

indicated in Fig. 3b, the cold downwellings on the sides of the plate are pushed aside as the sub-plate mantle becomes heated. In this heated state the plate is under significant tension ($\sim 70 \text{ MPa}$). The cold downwellings can no longer keep the plate trapped and it rapidly moves off the hot upwelling, relaxing the high tensional stress. A peak velocity of $\sim 7 \text{ cm yr}^{-1}$ is reached (Fig. 3d) at about the same time that σ_{xx} is reduced to zero; the sub-plate temperature also peaks ($\Delta T_b \sim 250^\circ\text{C}$) just before maximum plate velocity is achieved. In Fig. 3, a vertical arrow indicates the closest downwelling off to the right of the original plate position; Fig. 3b-e shows the plate moving toward this downwelling and coming to rest over it. As the plate settles over the downwelling, the stress moves into a compressive state (Fig. 3e) of more than 50 MPa. The gross topography of the plate can be estimated from the average horizontal vertical stress. Initially, the plate is topographically high and reaches a peak elevation when σ_{xx} is at a maximum (Fig. 3b), but rapidly subsides during rapid plate motion; the total (peak to peak) variation in topography over time is $\sim 3 \text{ km}$. The part of the fluid vacated by the plate is significantly overheated with a thin thermal boundary layer on the top surface (Fig. 3d). This is an unstable state and the convecting fluid passes into a chaotic flow regime (Fig. 3f). Although this is a subject of active research¹⁴, benchmarks at different resolutions suggest that this chaos is adequately resolved with the 33 nodes in depth¹². The cycle starts in an organized state with regularly spaced convection cells (Fig. 3a) but ends in a disorganized state (Fig. 3f), although in both instances the average plate velocity is either zero or very low.

The cycle just described is general, but the period and relative length of the stationary interval to rapid-velocity interval is dependent on plate size: the period between bursts of high velocity decreases with increasing plate width, until by $w_c \approx 4$ the plate always has an appreciable velocity (Fig. 2). Apparently, for this relatively long plate, a large-scale horizontal temperature gradient always exists and continually pushes the plate along, although there are still short bursts of higher velocity. For

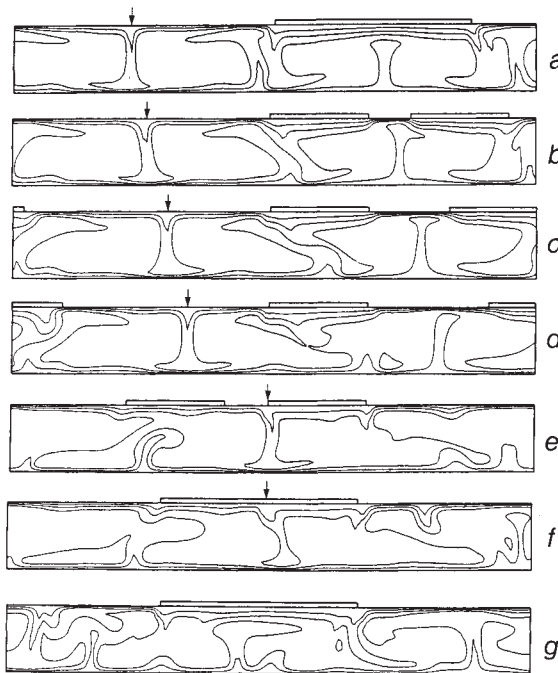


Fig. 5 Isotherms for a plate of $w_c = 3$ which breaks into two plates of $w_c = 1.5$ under a tensile yield stress, σ_{yld} , of 70 MPa. The arrow points toward a cold downwelling which the continental fragments move toward after break-up. The frame of reference is fixed with respect to the left corner of the original plate; the right plate in frame *b* moves with respect to the left plate, which is stationary in this reference frame.

$1 \leq w_c \leq 3$ the peak continental plate velocity is almost constant at $\sim 7\text{--}8 \text{ cm yr}^{-1}$, whereas for $w_c = 4$, the peak velocity is distinctly less, $\sim 4.5 \text{ cm yr}^{-1}$. The root mean square of the velocity histories of Fig. 2, however, are almost constant for all cases, including $x_c = 4$, at $\sim 1.5\text{--}2.0 \text{ cm yr}^{-1}$, with some suggestion of a peak between $w_c = 2$ and 3. For small widths, $w_c \ll 1$, the plate has no effect on the flow and aligns itself perfectly on top of a downwelling; if the flow is intrinsically steady, which is true for the Rayleigh number used in this study, then the plate velocity will be zero.

Breaking-healing plate

We now extend our analysis to the case of a long plate that can break under tension and heal upon collision; this model problem may contain most of the relevant physics involved in the breakup and assembly of supercontinents. As the starting state for the breaking-healing plate, an instant from a single plate run with $w_c = 3$ was used. When the plate breaks, each plate of the pair will have $w_c = 1.5$. The tensile yield stress, σ_{yld} , was set at 70 MPa, a reasonable value based on the failure of rocks studied in the laboratory¹⁵. Figure 5a is a few time steps before σ_{xx} reached σ_{yld} . If this plate could not yield under tension then the subplate temperature would have increased further, making $\sigma_{xx} > 70 \text{ MPa}$. In Fig. 5a, the plate is in a stable state with cold downwellings on either side and a hot upwelling in the centre. If more freedom were allowed in the break-up algorithm, the plate would still break in the centre because of symmetry. The frame of reference for Fig. 5 is with respect to the left margin of the original plate, such that the fluid in the box will flow by this stationary plate; this frame of reference is different from that of Fig. 3 which was in the mantle frame.

Immediately after the long plate breaks, the two smaller pieces start separating rapidly with a velocity of $\sim 7 \text{ cm yr}^{-1}$ by 4 Myr. By 18 Myr after break-up (Fig. 5b), the plates are separating at 10.5 cm yr^{-1} . Moreover, the hot upwelling extends to the surface between the two continental plates and forms a thin thermal boundary layer. The cold downwellings, originally on the two edges of the long plate, now point toward the region of hot upwelling (Fig. 5c). Both plates rapidly move toward the cold downwellings on the other side of the box (marked with an arrow, Fig. 5). As the frame of reference is with respect to the left corner of the original long plate (now the centre plate),

this cold downwelling is seen to move toward the centre plate. The two plates collide $\sim 150 \text{ Myr}$ after breakup. As in the case with the single plate, the area vacated by the continental plates is significantly overheated and undergoes frequent boundary layer instabilities which bring the fluid into a chaotic flow regime. After break-up, the two small plates spend most of the time under compressive horizontal stress. Immediately after assembly, the new long plate is under compression and positioned over the cold and downward flowing part of the fluid (Fig. 5f). In Fig. 5g, nearly 450 Myr after break-up, a new upwelling has developed under the plate. An estimate of the overall topography variation of the plate can be deduced from σ_{zz} . The peak-to-peak variation in topography is $\sim 3.5 \text{ km}$ from a high elevation before and during break-up to a minimum elevation following assembly.

A slightly different model has been run for a significantly longer period of time; in Fig. 6 the horizontal axis is distance across the top of the convecting fluid, and the vertical axis is time. Continental plates are denoted in black. This model differs from the case in Fig. 5 in that a smaller yield stress is used ($\sigma_{yld} = 50 \text{ MPa}$) but the aggregated plate has the same width ($w_c = 3$). The run starts with the plate in the aggregated state and remains that way as it undergoes slow lateral translation at $\sim 1 \text{ cm yr}^{-1}$; the plate breaks at 700 Myr. The fragments move apart and assemble again after 350 Myr, having reached velocities of up to 7 cm yr^{-1} . At the 700 Myr break-up, the flow exhibited a significant degree of order with only one major downwelling opposite the supercontinent, so that reassembly was relatively rapid; this is similar to the case shown in Fig. 5 which had a higher yield stress. At 1,100 Myr the reassembled continent remains stationary and insulates the mantle, so that after about another 400 Myr the continent breaks up again. After the 1,500-Myr break-up, the flow is quite chaotic and each continental fragment comes to rest over a different cold downwelling (at 1,900 Myr); the chaos apparently inhibits the fragments from reassembly for at least the next 2,000 Myr. Clearly, break-up is an almost certain occurrence if a supercontinent is assembled but assembly of small fragments over the same cold downwelling can be a long term process controlled by the number of downwellings and the degree of chaos in the flow. During the 3,500 Myr-model run, two break-ups and one assembly occurred. The continents were disassembled or fragmented for 70% of the time.

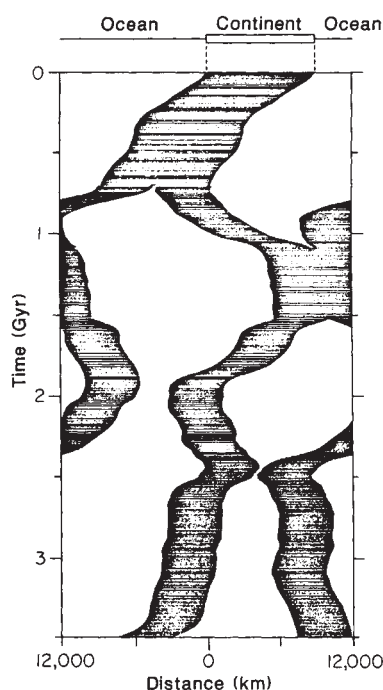


Fig. 6 Geometry of the continental plate as a function of time. The continental plates are denoted in black. The parameters used to scale the results are given in Fig. 4 legend.

Discussion

Despite the apparent simplicity of the models, the physics just outlined is in overall agreement with the geological record, especially during the last 200 Myr since the breakup of Pangaea. The Pangaea supercontinent, in its completely assembled state, was positioned over what is now the Atlantic-African geoid high³. Unfortunately we have no direct evidence of the form of the geoid 200 Myr ago, but today the Atlantic-African geoid high (like the Pacific geoid high) is strongly correlated with the location of present-day hot spots³⁻⁵. If the geoid high was correlated with hotspots 200 Myr ago, then it must have been approximately centred over Africa, as Africa has moved only slowly with respect to hotspots since breakup. The geoid high and hotspots are also correlated with slow seismic velocities in the lower mantle at spherical harmonic degree two, where most of the power is for both the geoid and the hot-spot locations^{6,7}. These observations are consistent with a lower mantle that was hotter than average under Pangaea before break-up, exactly as predicted by the numerical models. Since the supercontinent

break-up, the continental fragments North and South America, Antarctica, Australia and India have been moving towards geoid lows¹⁶.

The present-day pattern of stress in continents is also consistent with the state of stress in the modelled plates. The models predict that after break-up, when continental plates are settling in over cold downwellings, plates should be in a compressive state. In general most continental plates are in a state of compression¹⁷⁻¹⁹; the exception is Africa, which has significant portions under tension¹⁷. This anomalous pattern for Africa is consistent with its lying over hot mantle, as indicated by the geoid, hotspot locations and lower-mantle seismic tomography. Following the break-up of Pangaea, the continents experienced major transgressions in the late Cretaceous²⁰, a transgression being consistent with continental subsidence, a rise in sea level, or a combination of both²⁰. The numerical models predict a rapid rise in relative sea level or relative subsidence after continental break-up.

Within uncertainty, the time scales and plate velocity are also consistent with observations. Just after the Pangaea break-up, the average continental velocity was $\sim 4 \text{ cm yr}^{-1}$ (ref. 21) and there has been an apparent slowing down of the continents since then, present day velocity being $\sim 2 \text{ cm yr}^{-1}$ (refs 21, 22). This is consistent with the continents slowing down over cold-mantle downwellings. This 4-cm yr^{-1} observed velocity is somewhat lower than that of the model presented in Fig. 5 which had a separation velocity of 10 cm yr^{-1} . In general, it took a few hundred Myr for the modelled continents to settle in over cold downwellings after break-up. This reassembly time scale is dependent on both the aspect ratio of the box and the number of cold downwellings in the oceanic area before supercontinent break-up. For the case in Fig. 5, a high degree of order is present with only one cold downwelling opposite the long plate; this situation provided the minimum distance for fragments to traverse and reassemble, and hence a minimum time for the aggregation dispersal cycle. A maximum time required for the plates to move off hot mantle and toward cold downwellings is about 1,000 Myr; this is a maximum time because breaking plates move off sooner with a value determined somewhat by the tensile yield stress. A crude estimate of the time between assembled supercontinents, inferred from the numerical models, is in the range of a few hundred Myr up to two thousand Myr, with preferred values on the high end. Anderson³ suggests that the continents were dispersing 500–350 Myr ago (the state the continents are in today), so that the observed cycle time may be $\sim 500 \text{ Myr}$, consistent with the numerical results.

Discussions and valuable reviews were provided by Don Anderson and Bradford Hager. The work was supported by the NSF and a grant from the Ametek Corporation. Computing was carried out on the facilities of the San Diego Supercomputer Center.

Received 20 November 1987; accepted 7 March 1988.

1. Smith, A. G., Hurley, A. M. & Briden, J. C. *Phanerozoic Palaeocontinental World Maps* (Cambridge University Press, 1981).
2. Kerr, R. A. *Science* **230**, 1364–1367 (1985).
3. Anderson, D. L. *Nature* **297**, 391–393 (1982).
4. Crough, S. T. & Jurdy, D. M. *Earth planet. Sci. Lett.* **48**, 15–22 (1980).
5. Chase, C. G. *Nature* **292**, 464–468 (1979).
6. Hager, B. H., Clayton, R. W., Richards, M. A., Comer, R. P. & Dziewonski, A. M. *Nature* **313**, 541–545 (1985).
7. Hager, B. H. & Clayton, R. W. in *Mantle Convection* (ed. Peltier, W. R., in the press).
8. Hager, B. H. & O'Connell, R. J. *J. geophys. Res.* **86**, 4843–4867 (1981).
9. Davies, G. F. *Geophys. J. R. astr. Soc.* **84**, 153–183 (1986).
10. Olson, P. & Corcos, G. M. *Geophys. J. R. astr. Soc.* **62**, 195–219 (1980).
11. Schmeling, H. & Jacoby, W. R. *J. Geophys.* **50**, 89–100 (1981).

12. Hager, B. H. *et al. Los Alamos National Laboratory Technical Report* (in the press).
13. McKenzie, D. P., Roberts, J. M. & Weiss, N. O. *J. Fluid Mech.* **62**, 465–538 (1974).
14. Vincent, A. P. & Yuen, D. A. *University of Minnesota Supercomputer Institute Technical Report* **87/56**.
15. Brace, W. F. & Kohlstedt, D. L. *J. geophys. Res.* **85**, 6248–6252 (1980).
16. Anderson, D. L. *Science* **223**, 347–355 (1984).
17. Richardson, R. M., Solomon, S. C. & Sleep, N. H. *Rev. Geophys. Space Phys.* **17**, 981 (1979).
18. Zoback, M. L. & Zoback, M. J. *geophys. Res.* **85**, 6113–6156 (1980).
19. Lambeck, K., McQueen, H. W. S., Stephenson, R. A. & Denham, D. *Ann. Geophys.* **2**, 723–742 (1984).
20. Bond, G. *Geology* **4**, 557–560 (1976).
21. Schult, F. R. & Gordon, R. G. *J. geophys. Res.* **89**, 1789–1800 (1984).
22. Irving, E. *Nature* **270**, 304–309 (1977).
23. Turcotte, D. L. & Schubert, G. *Geodynamics* (Wiley, New York, 1982).

S-process krypton of variable isotopic composition in the Murchison meteorite

Ulrich Ott*, Friedrich Begemann*, Jongmann Yang†‡ & Samuel Epstein†

* Max-Planck Institut für Chemie, Saarstrasse 23, 6500 Mainz 1, FRG

† Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

Current theories on the origin of the chemical elements explain the abundance of medium-heavy and heavy nuclides to be due to the capture by pre-existing lighter nuclides of free neutrons on either a slow timescale (s-process) or a rapid timescale (r-process)¹⁻³. Experimental evidence in support of these theories comes from the analysis of carbonaceous chondrites. In acid-resistant residues of these meteorites a kind of xenon has been found, the isotopic composition of which matches almost perfectly that predicted for s-process xenon^{4,5}. We report data that allow us, for the first time, to derive with reasonable precision the full isotopic spectrum of s-process krypton as well. We show that this s-Kr in a residue from Murchison meteorite did not originate in one single s-process but rather is a mixture of contributions from stellar environments where the density of free neutrons was not the same. The astrophysical conditions under which this krypton has been produced were distinct from those that have been invoked to explain the Solar System s-process abundance. Similar to the ¹³C-rich carbon component in an aliquot of the same residue⁶, the s-process Kr from different astrophysical sites has retained its identity during the accumulation and subsequent history of the meteorite.

The measurements were performed on 7.38 mg of residue CFOc (see ref. 6 for details on the sample). Yang and Epstein have reported extreme enrichments in deuterium (up to 2,850%) and ¹³C (up to 508%) in the 700–900 °C combustion steps of their aliquot. Our sample was pre-degassed for 1 h at 350 °C under vacuum, then combusted in steps at 700, 800, 900 and 1,080 °C, by exposing it for 4 h to an oxygen pressure of ~10 mbar, and finally, after transfer to a high-temperature extraction system, pyrolysed for 30 min at 1,000 °C (to release newly adsorbed air) and at 1,800 °C. Approximately 95% of the total Kr and Xe was released in the 700 °C combustion step. Within the limits of error the isotopic composition of this fraction agrees with that of AVCC-Kr, Xe (ref. 7); the gases released in the 1,000 °C pyrolysis step, on the other hand, show the signature of adsorbed noble gases from the atmosphere.

The four remaining temperature steps all show clear evidence for a 'strange' component. Figure 1 shows the results for xenon. They indicate the presence of a Xe component with all essential features of s-process xenon: normalized to the abundance of

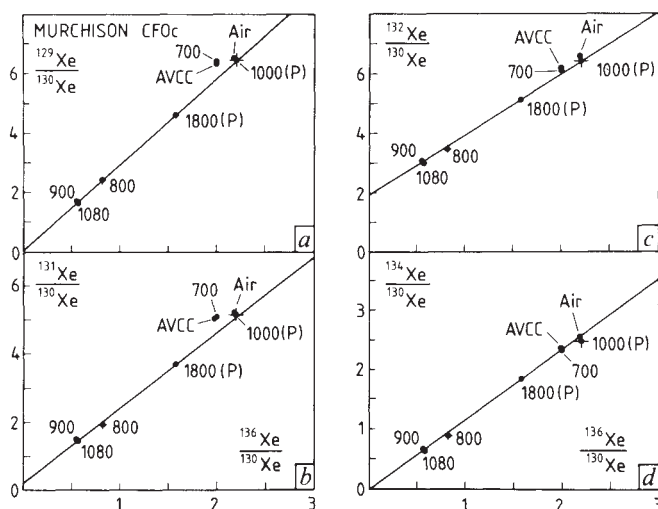


Fig. 1 Isotopic composition of Xe in Murchison residue CFOc. Data points form a linear array. The composition of s-process Xe is derived from the intersection with the ordinate (¹³⁶Xe = 0).

¹³⁰Xe the two lightest isotopes (¹²⁴Xe, ¹²⁶Xe; not shown) and the two heaviest ones (¹³⁴Xe, ¹³⁶Xe) are strongly depleted, and so are the two odd mass-number isotopes ¹²⁹Xe and ¹³¹Xe. The Xe released in the 900 °C and 1,080 °C combustion steps is the most s-Xe rich component as yet observed. Even more important, however, is the fact that the Xe in this residue appears to be a mixture of only two components. In three-isotope plots of the kind shown in Fig. 1, the data points from previous analyses^{8,9} are scattered all over the triangle determined by three poorly defined pure components. In this case, on the other hand, the data points fall along a single line, with the hypothetical s-xenon end member at the lower left. Under the usual assumption that this s-Xe is free of ¹³⁶Xe we find, from weighted fits to the data, its composition to be 128/129/130/131/132/134 = 0.46 ± 0.03/0.03 ± 0.11/1.02 ± 0.09/1.89 ± 0.10/−0.03 ± 0.05, which compares well with that derived previously^{8,9}. Note, however, that our approach also yields ¹³⁴Xe = 0 within the uncertainties (−0.03 ± 0.05; Fig. 1d), which had to be assumed in all previous efforts to derive the composition of s-Xe. Thus, although the errors of this composition are not significantly reduced as compared with the previous ones, the number of assumptions involved in its derivation are.

Tang *et al.*¹⁰ have commented upon the fact that the most s-xenon-rich fractions observed so far in different temperature steps of different residues from two different meteorites are all remarkably similar in their isotopic composition. We concur with their suggestion that this indicates the presence in carbonaceous chondrites of just such a component which was not produced by mixing *in situ* but rather existed before incorporation into its carrier(s). Actually, this is exactly what one expects to find if an s-process operates on matter which incorporates pre-existing heavy elements from a previous star generation.

‡ Present address: Department of Physics, College of Natural Sciences, Ewha Womans University, Seo Dae Moon-Gu, Seoul, Korea.

Table 1 Krypton released from Murchison residue CFOc at different temperatures during combustion and pyrolysis (P)

Extraction temperature	⁸² Kr	⁸⁰ Kr/ ⁸² Kr	⁸³ Kr/ ⁸² Kr	⁸⁴ Kr/ ⁸² Kr	⁸⁶ Kr/ ⁸² Kr
700 °C	36.1 ± 1.1	0.1942 ± 0.0122	1.0105 ± 0.0352	4.916 ± 0.123	1.526 ± 0.049
800 °C	0.211 ± 0.043	0.1704 ± 0.0214	0.7776 ± 0.0567	4.141 ± 0.156	1.241 ± 0.080
900 °C	0.366 ± 0.047	0.1250 ± 0.0176	0.6804 ± 0.0391	3.939 ± 0.139	1.382 ± 0.067
1,080 °C	0.516 ± 0.051	0.1276 ± 0.0104	0.6990 ± 0.0221	3.848 ± 0.094	1.281 ± 0.047
1,000 °C (P)	0.165 ± 0.013	0.1939 ± 0.0221	0.9542 ± 0.0473	4.827 ± 0.161	1.462 ± 0.075
1,800 °C (P)	0.441 ± 0.015	0.1703 ± 0.0120	0.8821 ± 0.0237	4.507 ± 0.080	1.536 ± 0.033

⁸²Kr abundance in units of 10^{−10} cm³ g^{−1} (STP)

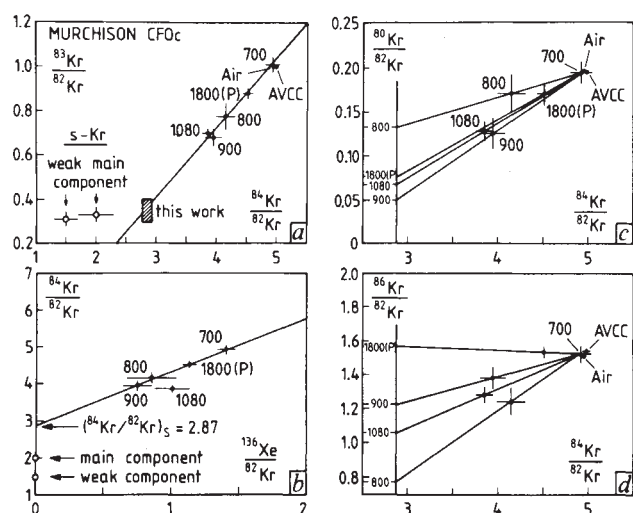


Fig. 2 Isotopic composition of Kr in Murchison residue CFOc. In a $^{83}\text{Kr}/^{82}\text{Kr}$ against $^{84}\text{Kr}/^{82}\text{Kr}$ plot the correlation line defined by the data misses the theoretical composition of s-Kr for main and weak s-process components (H. Beer, personal communication) (a). $^{84}\text{Kr}/^{82}\text{Kr}$ correlates with $^{136}\text{Xe}/^{82}\text{Kr}$ in different temperature steps. The intersection at $^{136}\text{Xe}/^{82}\text{Kr} = 0$ is interpreted as the $^{84}\text{Kr}/^{82}\text{Kr}$ ratio of s-Kr (b). $^{80}\text{Kr}/^{82}\text{Kr}$ (c) and $^{86}\text{Kr}/^{82}\text{Kr}$ (d) do not correlate with $^{84}\text{Kr}/^{82}\text{Kr}$. Apparently these ratios are not constant in the s-process component.

The krypton data are listed in Table 1 and shown in Fig. 2. In a plot of $^{83}\text{Kr}/^{82}\text{Kr}$ against $^{84}\text{Kr}/^{82}\text{Kr}$ (Fig. 2a) the data form again a linear array with the regression line passing close to AVCC-Kr⁷. It misses the theoretical points for both main-component and weak-component s-krypton (ref. 11 and H. Beer, personal communication) as they were invoked to explain the overall Solar System abundance of s-process nuclides^{12,13}. These two (independent) components differ in the amount of Fe seed material irradiated and in total neutron dose the seed nuclides were exposed to. Actually, the regression line for $^{83}\text{Kr}/^{82}\text{Kr}$ even passes to the right of the point given by the inverse ratios of the neutron-capture cross-sections ($\sigma_{82}/\sigma_{83} = 0.330 \pm 0.033$, $\sigma_{82}/\sigma_{84} = 2.26 \pm 0.31$)^{14,15} which are the theoretical maximum ratios for an s-process which starts out with iron-peak seed nuclei. The same incompatibility is seen from a plot of the $^{84}\text{Kr}/^{82}\text{Kr}$ isotope abundance ratio against the $^{136}\text{Xe}/^{82}\text{Kr}$ nuclide abundance ratio (Fig. 2b). With the exception of the 1,080 °C combustion step all points plot on a straight line. This suggests that in the different temperature steps no significant elemental fractionation occurred between Kr and Xe. The $^{84}\text{Kr}/^{82}\text{Kr}$ ratio of s-Kr can then be read off the ordinate intercept ($^{136}\text{Xe}_s = 0$) and yields $(^{84}\text{Kr}/^{82}\text{Kr})_s = 2.87 \pm 0.06$ (Table 2). From a similar graph of $^{83}\text{Kr}/^{82}\text{Kr}$ against $^{136}\text{Xe}/^{82}\text{Kr}$ we obtain $(^{83}\text{Kr}/^{82}\text{Kr})_s = 0.347 \pm 0.050$, which agrees well with the inverse ratio of the neutron-capture cross-sections of 0.330 ± 0.033 ^{14,15}.

For the two remaining ratios, $^{80}\text{Kr}/^{82}\text{Kr}$ and $^{86}\text{Kr}/^{82}\text{Kr}$, there is no obvious regular relation with $^{84}\text{Kr}/^{82}\text{Kr}$ (Fig. 2c, d). But it is not necessarily to be expected that all points fall on the same mixing line, with s-Kr of uniform composition as one end

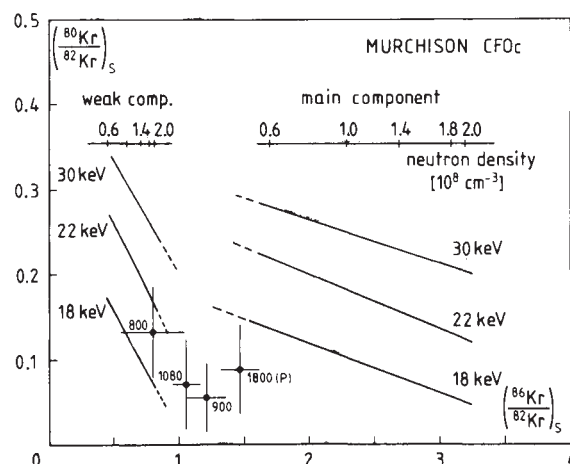


Fig. 3 Isotope abundance ratios $^{80}\text{Kr}/^{82}\text{Kr}$ against $^{86}\text{Kr}/^{82}\text{Kr}$ calculated for the weak and main component s-process as function of neutron density and temperature (H. Beer, personal communication). The range of parameters has been chosen such that superpositions of the two components reproduce the overall Solar System abundances of s-process nuclides. The experimental data points for s-Kr in Murchison residue CFOc are only reproduced with difficulty, if at all, indicating that this krypton was produced under conditions which were not representative for the genesis of average Solar System s-nuclides.

member. In contradistinction to all isotope ratios discussed so far, the $^{80}\text{Kr}/^{82}\text{Kr}$ and $^{86}\text{Kr}/^{82}\text{Kr}$ production ratios depend on details of the s-process, such as neutron density and temperature. This dependence happens because of the branchings in the s-process path at ^{79}Se and ^{85}Kr ^{11,14,16}. Because the 'terrestrial' half life of ^{85}Kr (10.76 yr) stays essentially the same under stellar conditions^{11,16}, at least up to $\sim 4 \times 10^8$ K, the $^{86}\text{Kr}/^{82}\text{Kr}$ production ratio depends only on neutron density n_n ; it increases with n_n because high neutron densities favour neutron capture over β -decay at ^{85}Kr . The half life of ^{79}Se ('terrestrial' 6.5×10^4 years), however, gets very much shorter under stellar conditions where the decay occurs predominantly via an allowed transition from a low-lying excited state at 96 keV^{14,16}. Consequently, this branching ratio depends not only on neutron density but on temperature as well. Qualitatively, high neutron densities result in low $^{80}\text{Kr}/^{82}\text{Kr}$ ratios, and high temperatures lead to high ratios.

We interpret the lack of a linear correlation between $^{80}\text{Kr}/^{82}\text{Kr}$ – $^{84}\text{Kr}/^{82}\text{Kr}$ and $^{86}\text{Kr}/^{82}\text{Kr}$ – $^{84}\text{Kr}/^{82}\text{Kr}$ to mean that the s-process krypton released from the same residue at different temperatures originated in stellar environments where neutron densities and/or temperature were not the same. This, at the same time, offers an explanation for the discrepant $(^{86}\text{Kr}/^{84}\text{Kr})_s$ ratios derived in previous analyses from the krypton released in different temperature steps from different residues^{8,9,17}. We derive the $(^{80}\text{Kr}/^{82}\text{Kr})_s$ and $(^{86}\text{Kr}/^{82}\text{Kr})_s$ ratios of our different temperature steps by the procedure indicated in Fig. 2c, d. For each temperature step the data point is connected with that of the 700 °C step and the $(^{80}\text{Kr}/^{82}\text{Kr})_s$ and $(^{86}\text{Kr}/^{82}\text{Kr})_s$ ratios are read off at the intercept of the extension of these straight lines

Table 2 Abundance of $^{82}\text{Kr}_s$ (in units of $10^{-10} \text{ cm}^3 \text{ g}^{-1}$ (STP) and isotopic composition of s-process Kr in Murchison CFOc

Temperature	$^{82}\text{Kr}_s$	$(^{80}\text{Kr}/^{82}\text{Kr})_s$	$(^{83}\text{Kr}/^{82}\text{Kr})_s$	$(^{84}\text{Kr}/^{82}\text{Kr})_s$	$(^{86}\text{Kr}/^{82}\text{Kr})_s$
800 °C	0.080 ± 0.023	0.132 ± 0.058	0.347 ± 0.050	2.874 ± 0.056	0.775 ± 0.260
900 °C	0.175 ± 0.034	0.049 ± 0.042	0.347 ± 0.050	2.874 ± 0.056	1.225 ± 0.147
1,080 °C	0.270 ± 0.036	0.067 ± 0.023	0.347 ± 0.050	2.874 ± 0.056	1.057 ± 0.098
1,800 °C (P)	0.088 ± 0.032	0.075 ± 0.064	0.347 ± 0.050	2.874 ± 0.056	1.574 ± 0.165

$(^{80}\text{Kr}/^{82}\text{Kr})_s$ and $(^{86}\text{Kr}/^{82}\text{Kr})_s$ show variations. The values for the individual temperature steps have been derived from Fig. 2c, d. (P), pyrolysis step.

with $^{84}\text{Kr}/^{82}\text{Kr} = 2.87$. This is equivalent to assuming that we are dealing with a two-component mixture of AVCC-like Kr and s-Kr with $^{84}\text{Kr}/^{82}\text{Kr} = 2.87$. The results are compiled in Table 2.

To show the information to be gained from this data, we have plotted in Fig. 3 the parameter-dependent ratios $(^{80}\text{Kr}/^{82}\text{Kr})_s$ against $(^{86}\text{Kr}/^{82}\text{Kr})_s$ calculated by H. Beer (personal communication) for different neutron densities and neutron temperatures. The two sets of curves pertain again to main-component and weak-component s-Kr, with the parameters characterizing the two components adjusted so as to reproduce the overall Solar System s-process abundances. With the exception of the 800 °C data point, our data fall outside the range for which the calculations have been performed. If the two intermediate points were a mixture of the two end members, or of two components lying even more to the left and right, respectively, then the same would be true for the right end member at least. Extrapolating the curves it is obvious that main-component Kr agrees with the Kr in our Murchison residue only for low neutron densities ($n_n < 0.4 \times 10^8 \text{ cm}^{-3}$) and low neutron temperatures ($T_n < 18 \text{ keV}$), and agreement with weak-component Kr is only marginally possible at high neutron densities of $n_n > 5 \times 10^8 \text{ cm}^{-3}$. The $(^{86}\text{Kr}/^{82}\text{Kr})_s$ ratio in the 1,800 °C pyrolysis step of 1.57 ± 0.17 (Table 2) is more than two standard deviations higher than the maximum possible ratio, for weak component s-Kr, of ~ 1.20 , however. It would seem that both components do not describe Murchison s-process Kr too well; but then, there is no reason why the particular s-process

products in a residue from Murchison carbonaceous chondrite should bear the same signature as average Solar System s-products.

We thank Dr H. Beer for allowing us to use the results of his calculations for the isotope abundance ratios of s-Kr produced under different stellar conditions. Work at the California Institute of Technology was supported by a NASA grant no. NAG 9-46. J.Y. gratefully acknowledges financial support by the Korea Science and Engineering Foundation and by the Third World Academy of Sciences.

Received 22 December 1987; accepted 22 February 1988.

1. Burbidge, E. M., Burbidge, E. R., Fowler, W. A. & Hoyle, F. *Rev. mod. Phys.* **29**, 547-650 (1957).
2. Fowler, W. A. *Science* **226**, 922-925 (1984).
3. Mathews, G. J. & Ward, R. A. *Rep. Prog. Phys.* **48**, 1371-1418 (1985).
4. Clayton, D. D. & Ward, R. A. *Astrophys. J.* **244**, 1000-1006 (1978).
5. Srinivasan, B. & Anders, E. *Science* **201**, 51-56 (1978).
6. Yang, J. & Epstein, S. *Nature* **311**, 544-547 (1984).
7. Marti, K. *Earth planet. Sci. Lett.* **3**, 243-248 (1967).
8. Alalerts, L., Lewis, R. S., Matsuda, J.-I. & Anders, E. *Geochim. cosmochim. Acta* **44**, 189-209 (1980).
9. Ott, U., Yang, J. & Epstein, S. *Meteoritics* **20**, 722-723 (1985).
10. Tang, M., Lewis, R. S. & Anders, E. *Meteoritics* **22**, 462-463 (1987).
11. Walter, G., Beer, H., Käppeler, F. & Penzhorn, R.-D. *Astr. Astrophys.* **155**, 247-255 (1986).
12. Käppeler, F. *et al. Astrophys. J.* **257**, 821-846 (1982).
13. Beer, H. in *Nucleosynthesis and its implications on Nuclear and Particle Physics* (eds Audouze, J. & Mathieu, N.) 263-270 (Reidel, Dordrecht, 1986).
14. Walter, G., Beer, H., Käppeler, F., Reffo, G. & Fabbri, F. *Astr. Astrophys.* **167**, 186-199 (1986).
15. Bao, Z. Y. & Käppeler, F. *At. Data nucl. Data Tables* **36**, 411-451 (1987).
16. Ward, R. A., Newman, M. J. & Clayton, D.D. *Astrophys. J. Suppl. Ser.* **31**, 33-59 (1976).
17. Matsuda, J.-I., Lewis, R. S. & Anders, E. *Astrophys. J.* **237**, L21-L23 (1980).

Formation of Xe-HL-enriched diamond grains in stellar environments

Uffe G. Jørgensen

Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

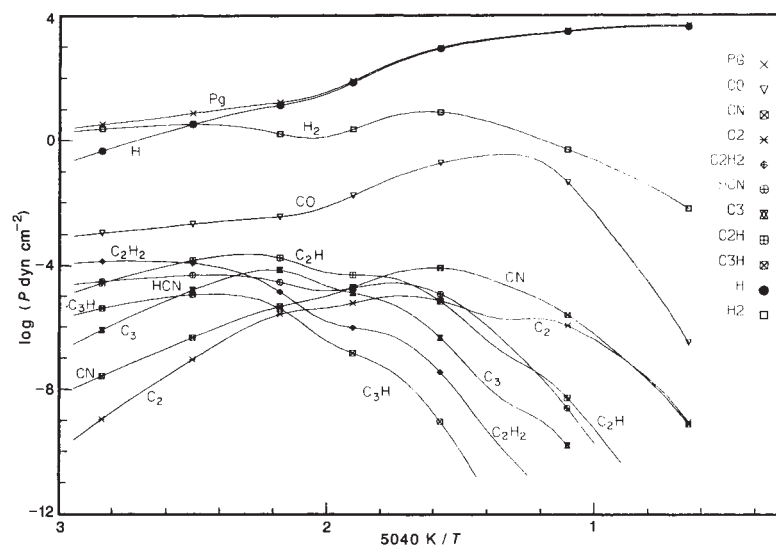
Recently Lewis *et al.*¹ have identified the C δ structure, previously found in C1, C2 and C3 carbonaceous chondrites, as diamond powder (that is, grains typically 50 Å in diameter). The C δ is enriched (compared to the atmosphere of the Sun) in heavy and light isotopes of Xe (Xe-HL) and therefore assumed to be of pre-solar origin. In laboratory experiments Roy² has described how small layers of diamonds can form at low pressure and at temperatures normally attributed to the grain-forming layers of cool stellar atmospheres ($\sim 1,000$ - $2,000 \text{ K}$). Lewis *et al.*¹ therefore suggest following the original idea of Clayton³, that the diamonds found in meteorites are formed in the upper atmosphere of red giant stars. Here it is shown that the diamonds must have formed in the type of red giants called carbon stars, and that Xe-HL and diamonds cannot have formed in the same star, but must be produced one in each component of a close binary system. The diamonds are produced in the smaller of the components, near the end of its evolution, the Xe-HL is produced when the compact remnant (a white dwarf) of the bigger component has accreted so much mass from the carbon star that it explodes as a supernova. This is consistent with the normal evolution of type I supernovae (SNI), and it is demonstrated that there were appreciably more carbon stars participating in SNI formation at the time of Solar System formation than today. Finally, the thermodynamic conditions are found under which the diamonds are produced, and it is explained why Xe-HL is found in diamonds but not in silicon carbide, and why s-process Xe is expected to be found in silicon carbide (C β).

There are two kinds of red giants; carbon stars have more carbon than oxygen in the atmosphere, and oxygen stars, mainly M-type giants, have more oxygen than carbon. Models predict

grains from oxygen stars to be of metal oxides (such as (Fe, Mg)SiO₄ and Al₂O₃; refs 4-6) and those from carbon stars to be of carbon compounds^{5,6} (silicon carbide and solid pure carbon). Diamonds have not yet been identified in stellar spectra, but the lack of detection can be ascribed to the fact that the predicted bands are diffuse and partly coincident with strong absorption from CO (at wavelengths 5 μ) and partly are in the ultraviolet ($\sim 7 \text{ eV}$) where the flux from carbon stars is negligible. The species C, C₂ as well as C₃ have been identified (along with a number of other carbon-containing gases) in the spectra of carbon stars, and C₃H is predicted to be abundant in cool carbon stars. All these molecules are expected to be important in the growth of solid carbon grains^{1,2}.

Figure 1 shows a calculation of the thermodynamic and molecular equilibrium conditions in the atmosphere of a cool carbon star of temperature $T_{\text{eff}} = 2,600 \text{ K}$, gravity given by $\log g = 0.0$ and abundance ratio $\log (\text{C}/\text{O}) = 0.13$ ($\text{H}/\text{C} \approx 1,000$), typical values for stars observed to have grain formation. The code used in the calculation is, with minor modifications, the one described in ref. 7. The partial pressure of H and H₂ constitutes most of the total pressure, whereas CO is the dominating carbon-containing molecule. Diatomic carbon-rich molecules dominate among the rest of the molecules in the deeper layers, whereas polyatomic molecules become more important in the upper layers (lower temperatures) where grains presumably begin to form. The model atmospheres presented here include opacity from the polyatomic molecules HCN, C₂H₂ and C₃, which is of crucial importance to the structure of the upper layers of model atmospheres⁸. Inclusion of all these three molecules in the opacity calculation give model atmospheres that are in much better agreement with observations than are models without⁹. The inclusion of polyatomic opacity causes the atmosphere to expand which brings the pressure in the upper layers down by a factor of $\sim 10^3$ compared to models without, allowing C₃H to be one of the dominant species at low temperatures for the model in Fig. 1. This molecule is fundamental to diamond formation¹, but is absent in older model atmosphere calculations¹. The present model is not calculated to temperatures lower than those shown in Fig. 1, because grain formation itself is then expected to alter the structure. The ratio

Fig. 1 The total gas pressure (P_g) and the partial pressure of atomic hydrogen and various molecules calculated for a stellar atmosphere of $T_{\text{eff}} = 2,600$ K, $\log g = 0.0$ and elemental composition as in H II regions in LMC, except for carbon which is raised to make $C/O = 1.3$



of H_2 to H was in various experiments², found to be an important parameter for diamond formation on films. It is seen in Fig. 1 (and this holds true for calculations with slightly different choices of parameters) that the partial pressure of H_2 is, in the grain-forming layers, everywhere bigger than or similar to that of free hydrogen, and increases over the partial pressure of H outward in the atmosphere. The carbon star atmosphere presented in Fig. 1 is, in contrast to oxygen-rich atmospheres, in qualitative agreement with the conditions needed in the laboratory to produce diamond grains². Quantitatively, the calculated model atmospheres may give a hint to laboratory experiments concerning the precise conditions whereby nature produces micro-diamonds.

Carbon stars form from intermediate mass stars ($1M_{\odot} \leq M \leq 8M_{\odot}$ the precise limits depending on metallicity and other parameters^{10,11}). Stars in this mass interval develop a degenerate core, and during most of the asymptotic giant branch (AGB) evolution get their energy from hydrogen shell-burning. For shorter periods the main energy will come from flashes of helium shell-burning that will, at the same time, produce the carbon (^{12}C) that is later dredged up to make the star a carbon star. These stars end their evolution as white dwarfs.

Stars above $8M_{\odot}$, on the other hand, will ignite carbon non-degenerately in the core and avoid the development of helium shell flashes, and hence remain oxygen rich in the atmosphere. Only such stars can evolve to the traditional supernova explosion of type II¹² (a single star supernova explosion). Some years ago a theory of a so-called carbon-detonation supernova of stars of mass less than $8M_{\odot}$ was attempted, and was the basis for Clayton's^{3,13} idea of $C\delta$ and Xe-HL production in the same star, but the theory is now considered unlikely.¹² The same star thus cannot produce the diamonds and the Xe-HL found inside the diamonds. The only likely way of having the diamonds produced in one star and the Xe-HL in another is that the two stars belong to the same double star system.

Imagine now two stars of such mass that both can develop into carbon stars. When the smaller of the two becomes a carbon star, the other is already a white dwarf, because higher-mass stars evolve faster. If the two stars are close enough, mass from the carbon star will flow over to the white dwarf and if the white dwarf thus eventually exceeds the Chandrasekhar limit it will blow up as a type I supernova (SNI). Mass transfer to white dwarfs is the standard explanation for SNI. Even if the stars are not close enough to fill the Roche lobe, mass transfer will occur because of the huge stellar wind, part of which will be captured by the white dwarf. The model suggested here for the formation of the $C\delta$ found in carbonaceous chondrites is that the diamonds are produced in the carbon star under the conditions sketched in Fig. 1, and when mass transfer from the carbon

star forces the companion white dwarf to explode, Xe-HL will be produced by the supernova. The Xe-HL will leave the supernova with speeds of $\sim 10^3$ km s⁻¹ and will be trapped in the accreting grains by ion implantation as it was in the single star model¹³ for ion implantation in circumstellar grains. Double stars are very common in the Galaxy, and carbon stars definitely participate in mass transfer in binary systems. There are indications that a whole class of stars (Ba stars) has been created by mass transfer from carbon stars to oxygen stars in close binary systems¹⁴.

The deeper layers of a pre-supernova star may contain³ a carbon-rich zone with $C/O > 2$, and $^{13}C/^{12}C = 0$, and grains condensed from this zone may be enriched with Xe-H produced by neutrons from the $^{22}Ne(\alpha, n)$ reaction in the supernova shock front. It is nevertheless unlikely that grains produced in this way could have avoided inclusion of the much more reactive ^{129}I . If $C\delta$ is produced by this model it is therefore necessary that some mechanism must later remove the ^{129}I and its decay product before the $C\delta$ is included in the meteoritic material. For a more detailed description of this problem, see Clayton¹³. The ^{129}I decays to ^{129}Xe , which is not observed in excess in meteoritic material¹⁵. For this reason it is also improbable that the diamonds could have formed inside the pre-solar nebula, in some local inhomogeneity of newly included supernova gas.

Carbon stars constitute only a very small fraction¹⁶ of the red giant population in the solar neighbourhood, but they are the most luminous AGB stars, and therefore the stars with greatest mass loss. Furthermore, low mass ($1-2M_{\odot}$) carbon stars were much more common before Solar System formation. Rather surprisingly, these facts lead to the conclusion that as much as $\sim 90\%$ of the material that potential SNI progenitors might have captured before the Solar System formation must have come from carbon star red giants. The arguments are fairly simple and can be described schematically as follows.

Recently Strömberg¹⁷ has published a list of the metallicity of F-type dwarfs as a function of their age, based on photometry by Olsen¹⁸. The conclusion is that metallicity in the Galaxy increased very rapidly during the first $\sim 10^9$ yr to a value near $[Fe/H] = -0.8$, and then grew more slowly to $\sim$ solar value 6×10^9 yr ago. F-type dwarfs $\sim 12 \times 10^8$ yr old have $[Fe/H]$ values similar to H II regions of the Large Magellanic Cloud (LMC)¹⁹. Assuming that the chemical evolution of LMC was qualitatively similar to the evolution of the Galaxy, most carbon stars seen in LMC today ($M > 1M_{\odot}$) must have approximately the same metallicity as the H II regions. The logarithm of the number of carbon stars per unit mass of galactic material is approximately a linearly decreasing function of metallicity in (the H II regions of) a given galaxy,²⁰ and the same is true for the ratio of carbon stars to late M-type giants²¹. The ratio of

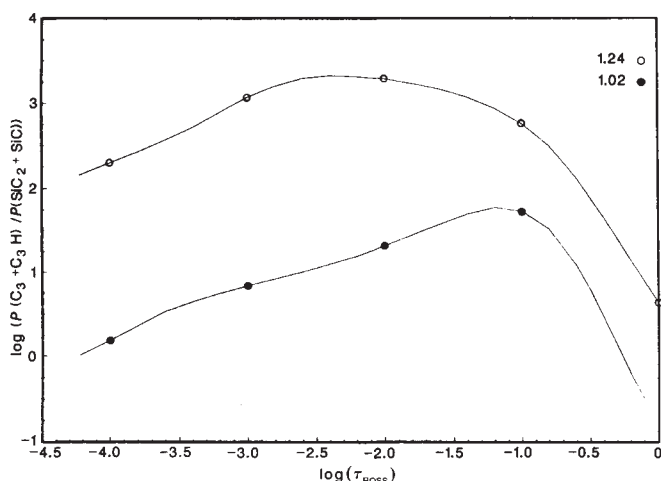


Fig. 2 The ratio, $P(C_3 + C_3H)/P(SiC_2 + SiC)$, of molecular partial pressure in two model atmospheres with $C/O = 1.24$ and 1.02 (both with $T_{eff} = 2,800$ K and $\log g = 0.0$). Horizontal axis is the logarithm of the Rossland mean optical depth (opaqueness) in the atmosphere. Grains are formed in the upper part of the atmosphere (left part of the diagram) and outward (leftward).

carbon stars to M-type giants later than M5 is 0.8 in LMC²¹. These arguments taken together mean that all smaller carbon stars in the solar neighbourhood at the time of the Solar System formation will have had the same metallicity that carbon stars (and other AGB stars) in LMC have today (stars of $M \approx 1-2 M_\odot$ will have total evolutionary time scales $\approx 7 \times 10^9$ yr, so those stars that were born $\approx 12 \times 10^9$ yr ago, when the Galaxy had the metallicity of the LMC today, would be in the final stages (AGB) of evolution when the solar system was born). The ratio, C/M , of carbon to M-type stars must therefore have been about 0.8 for the group of low mass AGB stars. Bigger stars will have been born out of interstellar material with metallicity nearer to the solar value, and they would therefore have created lower C/M than the low mass group. For simplicity, I will divide the red giant stars at the time of Solar System formation in only two populations: smaller stars with $M \approx 1-2 M_\odot$ and $C/M \approx 0.8$, and bigger stars with $M \approx 2-8 M_\odot$ and C/M of a few per cent.

The most widely accepted estimate of the number, $N(M_i)$, of stars with mass M_i is the Salpeter initial mass function (IMF), $N(M_i) = M_i^{-2.35}$. With this IMF, two thirds of the red giants will belong to the low mass group and one third to the high mass group. Assuming a mass loss rate, $M \propto LRM^{-1}$ (L , R are luminosity and radius) for oxygen-rich stars²² to hold also for carbon stars, we can estimate the ratio, M_c/M_0 , of the total mass loss, M_c , during the carbon stage, to the total mass loss, M_0 , during the oxygen star stage of AGB. Assuming that approximately equal periods in $\log L$ are lived during equal time intervals²³, $T_{eff} = \text{constant}$ during AGB evolution, and that the difference in magnitude between beginning and end of the AGB evolution of a giant star is $\Delta M \approx 5$ mag, we get, for a star in the low mass group, 300 times greater mass loss when the star is a carbon star (upper 80% of AGB) than when it is an oxygen star (lower 20% of AGB). This number may be somewhat too high, because it assumes that the ratio $M/C = 0.8$ measured in LMC is the ratio $M(AGB)/C$ of oxygen AGB stars to carbon star number (lifetime), which may not be exactly the case. Nevertheless, even assuming $M(AGB)/C = 0.2$, which may be in better agreement with theoretical predictions¹⁰, 75% of the mass loss from a star in the lower mass group should occur during the carbon star phase, and only 25% during the oxygen-rich AGB phase. The predictions of the lifetime of low-mass carbon stars in ref. 10 should be adjusted upward a little if the more recent results from ref. 11 are incorporated in the calculations of ref. 10, and in reality the mass loss rate, $\dot{M}_c$, for carbon stars is probably greater than the rate, $\dot{M}_0$ for oxygen rich stars,

among other reasons because of the higher temperature of carbon grain formation⁶. For these reasons I persist in assuming that virtually all the mass lost by stars in the low mass group occurs during the carbon star phase. Assuming $M_c/M_0 \approx 3$ and $C/M \approx 0.02$ for the high mass group of red giants gives $M_c/M_0 \approx 0.3$; that is, 30% of the mass loss from the high mass group is from carbon stars even if carbon stars contribute only 2% of the high mass AGB population.

Integrating numerically $IMF \times \dot{M}$ over the whole mass interval in the two groups, assuming the AGB lifetime to be independent of mass, gives the ratio of the total mass loss from all stars in the $1-2 M_\odot$ group to that of the $2-8 M_\odot$ group to be 1.5 (still assuming $M_c/M_0 \approx 3$). The total mass injection to the interstellar medium from AGB stars, near the time of solar system formation, was thus 60% from low mass AGB and 40% from high mass AGB stars. From the low mass stars all material was lost during the carbon star phase, but for the high mass stars 30% was lost during the carbon star phase and 70% during the oxygen rich phase. The ratio of material lost from all stars during their carbon phase to material lost from all AGB stars during their oxygen-rich phase was thus

$$\left(\frac{1.5}{2.5} 100\% + \frac{1}{2.5} 30\% \right) : \frac{1}{2.5} 70\% \approx 70\% : 30\%.$$

Furthermore, it is more likely for a low mass than a high mass AGB star to have a white dwarf companion, the relative frequency being $\sim 4:1$ for typical representatives ($M \approx 1.4 M_\odot$ and $3 M_\odot$), of both mass groups. The ratio of the total amounts of material transferred to white dwarf supernova progenitors from carbon stars to that transferred from oxygen AGB stars is then

$$\left(\frac{4}{5} \frac{1.5}{2.5} + \frac{1}{5} \frac{1}{2.5} 0.3 \right) : \left(\frac{1}{5} \frac{1}{2.5} 0.7 \right) \approx 90\% : 10\%.$$

These arguments are, of course, very qualitative in some places, but the conclusion is nevertheless obvious: in spite of the fact that carbon stars today constitute only a very small fraction of the total red giant population, a supernova I progenitor, before solar system formation, will have most likely accreted matter in the form of grains produced in a carbon star atmosphere. The last question we will examine is why it was diamonds and not silicon carbide that carried Xe-HL, when both pure carbon grains and grains of silicon carbide are expected to form in carbon rich atmospheres.

As a carbon star evolves, ^{12}C will be produced and mixed to the surface in increasing quantity, and the C/O ratio will increase. According to a recent investigation²⁴ of 30 of the brightest carbon stars (disregarding 6 stars with $^{12}C/^{13}C < 20$ which may originate in a distinct way, perhaps by CNO cycling), two distinct groups can be discerned, one having 7 stars with $^{12}C/^{13}C > 65$, and a mean $\langle C/O \rangle = 1.24$, and the other having 17 stars with $^{12}C/^{13}C < 65$ ($\approx 42 \pm 20$, for reasons mentioned below) and $\langle C/O \rangle = 1.07$. It is obvious from the $\langle C/O \rangle$ values that the first group is more evolved. Fig. 2 shows the ratio of partial pressures of molecular $C_3 + C_3H$ and molecular $SiC_2 + SiC$ for model atmospheres of a typical evolved carbon star ($C/O = 1.24$) and of one with smaller mass loss ($C/O = 1.02$), respectively. The molecules C_3 and C_3H are supposed to be progenitors for the solid carbon grains; SiC_2 and SiC are supposed to be the molecular progenitors for the silicon carbide grains. The big difference in the ratio of these molecules in the two models (a factor of more than 100 in the upper, grain forming layers) clearly indicates that the graphite and diamonds will be preferentially produced at the end of carbon star evolution (where C/O and $^{12}C/^{13}C$ are high and the probability for having transferred enough mass to the white dwarf to create a supernova explosion is biggest). The silicon carbide, on the other hand, is seen to be produced preferentially earlier in carbon star evolution when C/O and $^{12}C/^{13}C$ are low. The ratio $^{12}C/^{13}C$ has been measured²⁵ in diamonds ($C\delta$), isolated from

carbonaceous chondrites, to be $^{12}\text{C}/^{13}\text{C} \approx 92$, whereas the isotopic ratio for the C β structure, containing Xe in s-processed ratio (as produced in AGB stars themselves), has been measured²⁵ to be $^{12}\text{C}/^{13}\text{C} \approx 42$.

An identification of C β structures as SiC grains in the Murray meteorite has now been announced^{26,27} and the details in the announced analysis are in full agreement with the theory presented above for the formation of C δ and C β inclusions in carbonaceous chondrites. Enlightening discussions with H. Nørsgård and E. H. Olsen are greatly acknowledged. This work has been supported by the Danish Natural Science Research Council.

Received 3 December 1987; accepted 27 January 1988.

- Lewis, R. S., Ming, T., Wacker, J. F., Anders, E. & Steel, E. *Nature* **326**, 160–162 (1987).
- Roy, R. *Nature* **325**, 17–18 (1987).
- Clayton, D. D. *Astrophys. J.* **199**, 765–769 (1975).
- Schmid-Burgk, J. & Scholz, M. in *Physical Processes in Red Giants* (eds Iben, I. Jr. & Renzini, A.) 341–346 (Reidel, Dordrecht, 1981).
- Draine, B. T. in *Physical Processes in Red Giants* (eds Iben, I. Jr. & Renzini, A.) 317–333 (Reidel, Dordrecht, 1981).
- Salpeter, E. E. *Rev. Mod. Phys.* **46**, 433–436 (1974).
- Eriksson, K., Gustafsson, B., Jørgensen, U. G. & Nordlund, Å. *Astr. Astrophys.* **132**, 37–44 (1984).
- Jørgensen, U. G., Almlöf, J. & Siegbahn, P. E. M. *J. Chem. Phys.* (submitted).
- Jørgensen, U. G. *Thesis*, Copenhagen Univ., 1987.
- Renzini, A. & Voli, M. *Astr. Astrophys.* **94**, 175–193 (1981).
- Bertelli, G., Bressan, A. G. & Chiosi, C. *Astr. Astrophys.* **150**, 33–52 (1985).
- Woosley, S. E. in *Nucleosynthesis and Chemical Evolution* (eds Hauck, B., Maeder, A. & Meynet, G.) 1–195 (Geneva Observatory, 1986).
- Clayton, D. D. *Proc. Lunar Planet. Sci.* **12B**, 1781–1802 (Pergamon Press, Oxford, 1981).
- Lambert, D. L. in *Cool Stars with Excesses of Heavy Elements* (eds Jaschek, M. & Keenan, P. C.) 191–223 (Reidel, Dordrecht, 1985).
- Lewis, R. S. & Anders, E. *Astrophys. J.* **247**, 1122–1124 (1981).
- Miller, G. E. & Scalo, J. M. *Astrophys. J.* **263**, 259–268 (1982).
- Strömberg, B. in *The Galaxy* (eds Gilmore, G. & Carswell, R.) (Reidel, Dordrecht, in the press).
- Olsen, E. H. *Ast. Astrophys. Suppl.* **57**, 443–466 (1984).
- Peimbert, M. & Torres-Peimbert, S. *Astrophys. J.* **193**, 327–333 (1974).
- Richer, H. B. & Westerlund, B. E. *Astrophys. J.* **264**, 114–125 (1983).
- Cook, K. H., Aaronson, M. & Norris, J. *Astrophys. J.* **305**, 634–644 (1986).
- Reimers, D. *Mem. Soc. Roy. Sci. Liege. Ser. 6* **8**, 369–382 (1975).
- Iben, I. Jr. & Renzini, A. *Ah. Rev. Astr. Astrophys.* **21**, 271–342 (1983).
- Lambert, D. L., Gustafsson, B., Eriksson, K. & Hinkle, K. H. *Astrophys. J. Suppl.* **62**, 373–425 (1986).
- Lewis, R. S. & Anders, E. *Scient. Am.* **249**, 54–66 (1983).
- Bernatowicz, T. et al. *Nature* **330**, 728–730 (1987).
- Zinner, E., Ming, T. & Anders, E. *Nature* **330**, 730–732 (1987).

Gas and dust jets in the inner coma of comet Halley

Cristiano B. Cosmovici*, Gottfried Schwarz†, Wing-Huen Ip‡ & Peter Mack§

* Istituto di Fisica dello Spazio Interplanetario, CNR, 00044 Frascati, Italy

† DFVLR, Hauptabteilung Angewandte Datentechnik, 8031 Oberpfaffenhofen, FRG

‡ Max-Planck-Institut für Aeronomie, 3411 Katlenburg-Lindau, FRG

§ South African Astronomical Observatory, P.O. Box 9, Observatory 7935, South Africa

Cyanogen (CN) jets were discovered in comet Halley by applying an image processing ring masking technique to ground-based observations^{1,2}. We now report the detection of dust, C₂, C₃, CN and ([O I] + NH₂) jet features in comet Halley during the period of the spacecraft Giotto encounter (13–18 March 1986). With absolute flux measurements, careful continuum subtraction and an improved ring-masking technique, we have achieved an unambiguous separation of the gas and dust components of the cometary coma emissions. A comparison of the morphologies of these gas and dust jets provides important information about the anisotropic dynamical processes in the cometary coma and on the surface of comet Halley's nucleus.

The observations were carried out between 13–18 March 1986, around the time of the Giotto flyby of comet Halley on 13 March. The aim was to compare ground-based observations

Table 1 Instrumental and observational data

Observing site	Sutherland (SAAO)
Telescope	1.9-m Cassegrain f/6.5 reduced
Detector	Thinned RCA 53612 CCD
	360 × 512 pixels
Imaging area	318 × 510 pixels corresponding to: 2.97 × 4.76 arc min ²
	Average at comet: (120 × 10 ³ × 190 × 10 ³) km ²
Image resolution	0.56 arc s per pixel ≈ 370 km

Table 2 Filters used with the CCD camera system

Filters	Wavelength (Å)	FWHM* (Å)
International Halley Watch		
Continuum blue (BC)	3,650	80
CN	3,871	50
C ₃	4,060	70
CO ⁺	4,260	70
Mid continuum (MC)	4,845	65
C ₂	5,139	90
Max-Planck Institute		
H ₂ O ⁺	6,193	43
H ₂ O ⁺ continuum	6,253.5	18.7
[O I] + NH ₂	6,297	9.5
([O I] + NH ₂) continuum	6,248	8

* FWHM denotes full width at half maximum.

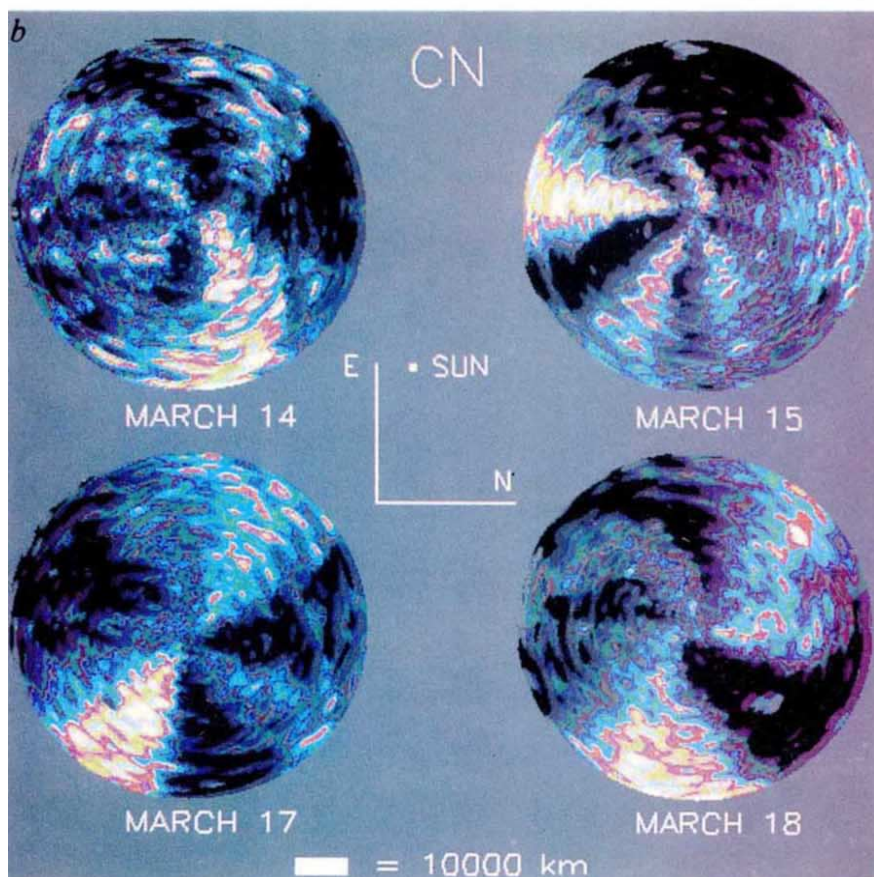
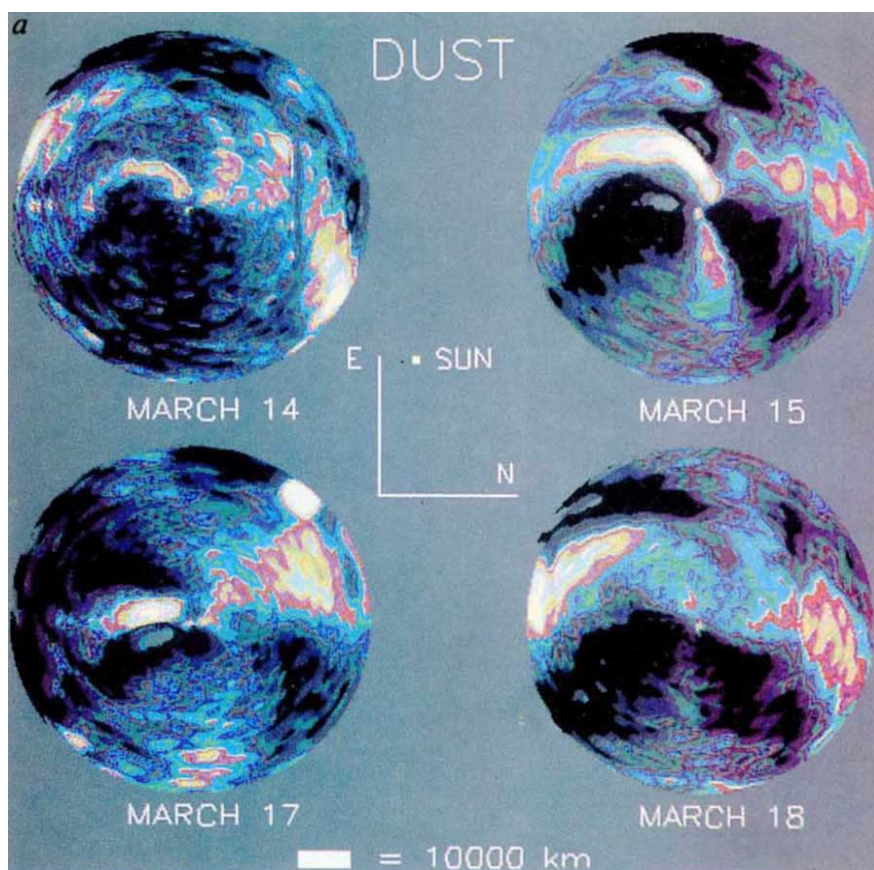
with close-up images taken by the Giotto camera^{3–5}. We used the charge-coupled device (CCD) camera system of the 1.9-m telescope at the South African Astronomical Observatory (SAAO), Sutherland, South Africa. Preliminary results, obtained with 23 different filters (including the filter set for the Halley Multicolour Camera of Giotto), have been reported^{3–6}. Further information about the instrumental and observational data and about the filters used is given in Tables 1 and 2. Table 3 summarizes the image data for the observations.

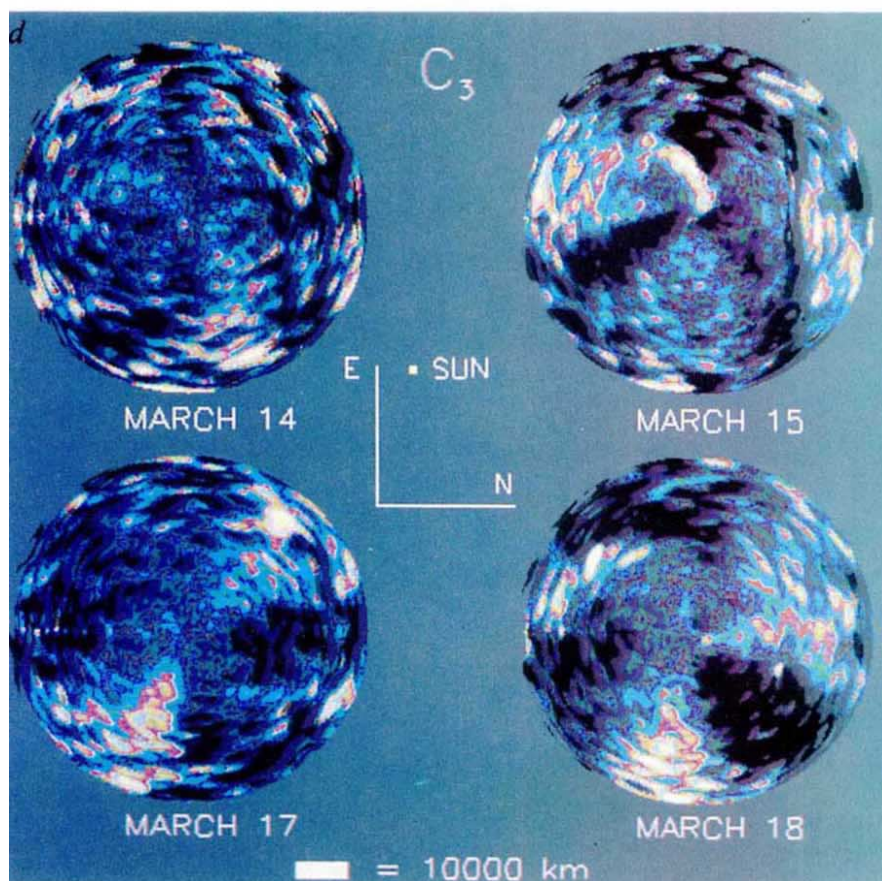
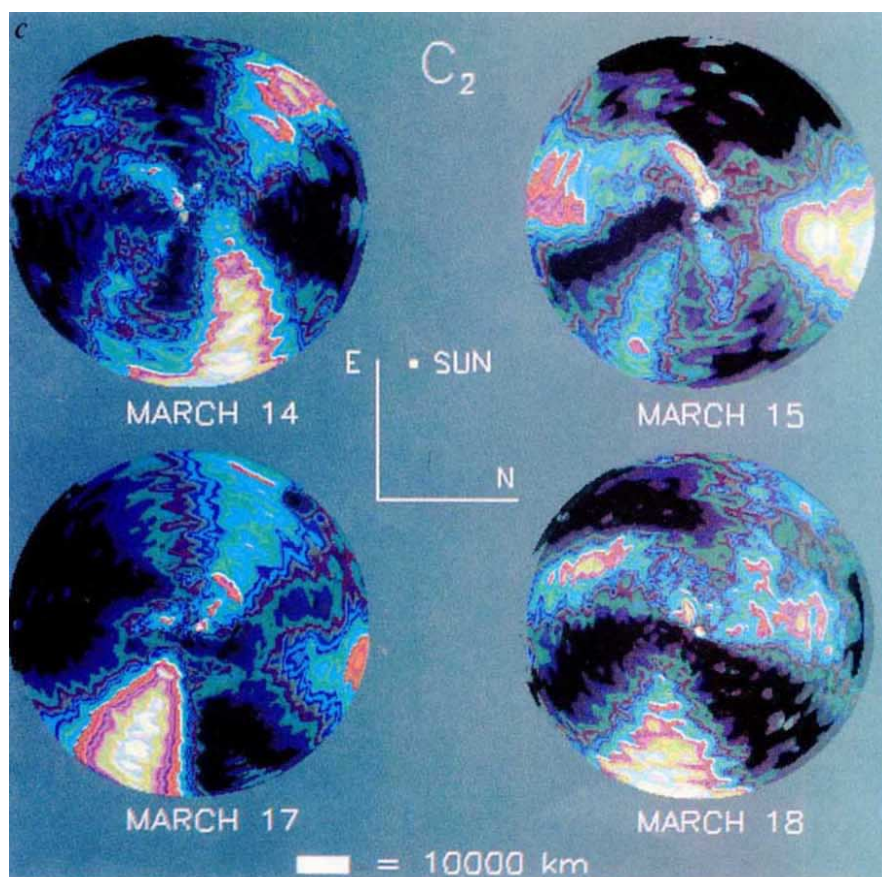
The technique for absolute photometric calibration as suggested by the International Halley Watch (IHW) for near-nucleus studies was used. As reported³, a locally adaptive ring-masking technique was used to enhance the jet features and it was shown that gas and dust jets can be resolved at different wavelengths in the cometary coma of comet Halley. The images obtained with narrow-band filters contain both gas emissions and continuum which stems from sunlight scattered on solid particles. After absolute calibration of all filters an accurate continuum subtraction was carried out by precisely co-aligning (that is superimposing with correct positioning) the gas and continuum filter images. In most cases, image edges had to be cut to remove non-overlapping areas and to eliminate blemishes caused by stars. For these reasons, the full images covering 120,000 × 190,000 km² could not be used for ring masking and the analysis was restricted to an area with a diameter of 70,000 km around the nucleus. Finally, the jet structures in the continuum subtracted images were enhanced by ring masking³.

In the case of H₂O⁺ and ([O I] + NH₂) emission we could use special narrow band filters with the continuum emission located very close to the gas emission wavelength (Table 2). Therefore, the background subtraction is quite accurate. In the case of the IHW filters the continuum wavelengths (3,650 and 4,845 Å) lie far from the C₃ and CO⁺ emissions, but as we had no jet structures in the continuum blue (3,650 Å) and the CO⁺ (4,260 Å) images showed only dust jet structures, we used these filters as continuum subtraction for CN (3,871 Å) and C₃ (4,060 Å), taking into account that the solar flux does not change appreciably in this wavelength region (Fig. 2, ref. 3). The 4,845 Å continuum filter was only used for subtraction from C₂ (5,139 Å). In the following, we concentrate on the images taken

Table 3 Image data for observations during Giotto encounter

Filter	Date (March 1986)	Exposure time (s)
B.C.	14.099	30
CN	14.108	30
C ₃	14.101	30
CO ⁺	14.104	30
M.C.	14.103	30
C ₂	14.105	30
H ₂ O ⁺	14.144	300
H ₂ O ⁺ continuum	14.150	250
B.C.	15.092	60
CN	15.101	60
C ₃	15.094	60
CO ⁺	15.099	60
M.C.	15.096	60
C ₂	15.100	30
H ₂ O ⁺	15.106	60
H ₂ O ⁺ continuum	15.114	60
B.C.	17.048	120
CN	17.059	60
C ₃	17.051	60
CO ⁺	17.055	60
M.C.	17.053	60
C ₂	17.057	60
[O I]+NH ₂	17.153	300
([O I]+NH ₂) continuum	17.163	300
B.C.	18.058	60
CN	18.074	60
C ₃	18.059	60
CO ⁺	18.069	60
M.C.	18.067	60
C ₂	18.071	60
O I+NH ₂	18.087	180
([O I]+NH ₂) continuum	18.095	180

**Fig. 1** *a*, Dust jets enhanced by ring masking in a region of 70,000 km around the nucleus; *b*, CN jets; *c*, C₂ jets; *d*, C₃ jets; *e*, [O I]+NH₂ jets.



on 14, 15, 17 and 18 March using the IHW filter set and the special set of narrow-band filters for the H_2O^+ ions and the $([\text{O I}] + \text{NH}_2)$ emission, as detailed in Tables 2 and 3. The jet enhanced images obtained for dust, CN, C_2 , C_3 and $([\text{O I}] + \text{NH}_2)$ are shown in Fig. 1.

Dust. Dust jets have been observed extensively from the Giotto and Vega and from the ground³⁻¹⁰. Their formation is presumably caused by emissions of gas and dust in discrete regions on the surface of the cometary nucleus. The morphology and the time variation of these dust jet features have been used to deduce the rotation period(s) of comet Halley¹⁰⁻¹². Our measurements covering five nights in March 1986 are not sufficient to derive rotational periods but could be correlated with other ground-based observations, for example, with observations taken from Chile and Australia (having a time shift of 8 hours) and with space observations. Figure 1a shows the dust morphology during four nights using the continuum filter at 4,845 Å. The continuum at 6,253 Å shows the same features. Note that, as a rule, no jets are seen in the continuum at 3,650 Å.

On 15 March a strong outburst caused an increase of the peak total luminosity in the red filter image by a factor of four with respect to the 14 March observation⁴⁻⁵. This outburst can clearly be seen as the long bright spiral jet to the south-east with a length of ~40,000 km. On all images there are jet structures moving away from the nucleus in agreement with the view that these dust jets are generated at different time intervals. Dust jets are only observable in the red wavelength region because the sunlight is mainly scattered by grains with diameters $d > \lambda > 4,000$ Å, that is, the Mie and not the Rayleigh scattering is responsible for the visibility of the dust jets.

CN, C_2 , C_3 . A'Hearn *et al.*² already observed that the shape of the CN and C_2 jets is different from the continuum jets at 4,845 Å, but their measurements were still affected by the continuum band which is very strong in the case of the C_2 jets. From our measurements, obtained during a different period but with accurate continuum subtraction, we observe the following. First, CN, C_2 and C_3 emissions consist of jet structures different from the dust continuum; second, C_2 jets are very bright and well defined with general structures mostly similar to their CN counterparts; third, on 15 March, during the outburst, there appeared a strong conical CN jet to the south which was not observed with any other filter. We therefore conclude that C_2 and C_3 were formed in the same regions as CN in the gas jets on 14, 17 and 18 March. On 15 March a particular event led to the formation of CN only in a region south of the nucleus with an extent of more than 50,000 km. Several mechanisms have been proposed to explain the formation of CN jets. These include sublimation from the sub-micrometre organic CHON grains^{1,13} and emissions from the cometary nucleus surface with chemical inhomogeneities¹⁴. The DIDSY instrument on Giotto provided information on the cumulative flux distributions of the dust particles in the coma of comet Halley^{15,16}. The data were used to estimate that the total mass production rate of the sub-micrometre grains is approximately 10^3 g s^{-1} . This value is too low to account for the C_2 and CN jet production even if the radical jets contain no more than 10% of the total C_2 and CN emissions (for a relative abundance of 10^{-3} – 10^{-2} for C_2 and CN, the corresponding mass production rate is about 2×10^4 – $2 \times 10^5 \text{ g s}^{-1}$). This estimate thus casts doubt on the validity of the CHON grain origin (see also refs 2, 14).

The alternative idea of direct emission from the surface of the nucleus¹⁴ requires that the radiation-pressure acceleration should be similar for both C_2 and CN. If not, we expect a separation of the two components at large distances from the cometary nucleus (the same argument applies if we assume sub-micrometre grains as the parent particles of the C_2 and CN radicals as these small dust particles will be sorted out from the larger particles by the solar radiation pressure). In view of these issues, both proposals should be studied in more detail. One additional effect which also deserves consideration has to do

with the ion chemistry such as $\text{H}_2\text{O}^+ + \text{HCN} \rightarrow \text{H}_2\text{CN}^+ + \text{OH}$, followed by an electron dissociative recombination process $\text{H}_2\text{CN}^+ + e \rightarrow \text{CN} + \text{H}_2$. In this way, the presence of plasma structures (that is, ion streamers or envelopes) in the cometary ionosphere could also lead to the generation of jet-like features in CN and C_2 . Because of the lack of ion emission profiles, it is not possible to test this hypothesis at present. Furthermore, any increase of the CN and C_2 radicals in the tailward side due to the ion chemical processes in the plasma tail, would superimpose on the asymmetry caused by the solar radiation pressure acceleration. Because H_2CN^+ ions could be quite abundant in the inner ionosphere¹⁷, it is important that this possibility is not overlooked.

[O I] + NH_2 . One new feature in our CCD observations is the presence of anisotropic patterns in the brightness of the forbidden [O I] emission at 6,300 Å. Our filter bandwidth permits the detection of the NH_2 (0, 8, 0 → 0, 0, 0) band at 6,294.45, 6,295.83, 6,297.32, 6,298.62 Å. In the case of the global anisotropy, because of the short lifetimes of NH_3 and NH_2 molecules due to photodissociation, most of the NH_2 emission can be shown to be limited to a distance of $\sim 2 \times 10^4$ km from the nucleus. The contribution from NH_2 is therefore small beyond this region. As for the jets, whose brightness increment is typically just a few per cent of the general background, one cannot rule out the possibility that [O I] and NH_2 might be released from small dust streamers—even though this would imply different source particles for the [O I] + NH_2 jets as well as the CN and C_2 jets. Because the ring-masking technique subtracts atmospheric oxygen automatically from our measurements, we can assume that the $([\text{O I}] + \text{NH}_2)/\text{dust}$ ratio in a region of $\sim 70,000$ km around the nucleus is ~ 8 . We obtain $[\text{O I}]/\text{dust} = 5.3$, $\text{NH}_2/\text{dust} = 2.7$. In the following, we shall assume that most of the asymmetry and jet features seen in Fig. 1e are due to the [O I] atoms. We consider that the [O I] images are of great interest as they carry unique information on the distribution of the H_2O molecules in the coma. So far, except for a brief mention of the asymmetry in the OH brightness distribution as observed by IUE¹⁸ and the inference from the near-infrared spectrometer experiment on Vega of a sunward jet of the H_2O vapour¹⁹, no report has yet been published describing the global distribution of the water molecules.

Of the two photolytic processes^{20,21}, $\text{H}_2\text{O} + h\nu \rightarrow \text{H}_2 + \text{O}^1\text{D}$ and $\text{OH} + h\nu \rightarrow \text{H} + \text{O}^1\text{D}$, the first is more important in the region covered by our observations ($r < 10^5$ km). Thus, the [O I] observations on 17 and 18 March can be used to make the following deductions. First, the water production rate and hence the surface sublimation process are not as isotropic as might have been inferred from the neutral mass spectrometer observations at comet Halley^{22,23}. According to our present analysis, the degree of anisotropy is quite significant, that is, an amplitude of $\Delta I/I \approx 30\%$ at a radial distance of $r = 5,000$ km on 17 March and $\Delta I/I \approx 19\%$ on 18 March. This variation probably reflects the presence of temporal variations in the outgassing process from the nucleus. Second, the [O I] jets show a similar morphology to those of the dust particles only on 18 March, but the south-pointing jet is $\sim 6,000$ km broader. They are also separated from the CN and C_2 jets. On 17 March, in addition to a wide jet pointing to the left which is related to a similar dust feature, the [O I] emission has a strong increase in the sunward-pointing direction.

One aim of the imaging observations using the pair of filters with central wavelengths at 6,193 Å and 6,253 Å is to resolve the emission from the H_2O^+ ions in the coma. The bandwidth of the filters was very narrow in comparison with the IHW filters (see Table 2). Because of the very strong emission from the dust continuum within a radial distance of a few 10^4 km, however a determination of the morphologies of the ion emission is not yet feasible.

In summary, detailed image processing of the ground-based observations from South Africa taken during the Giotto encoun-

ter has revealed a wealth of information on the global and local structures of the gas and dust comae of comet Halley. While our results confirm and expand on the discovery of CN jets by A'Hearn *et al.*^{1,2}, the unambiguous detection of C₂, C₃ and ([O I] + NH₂) jets separated from dust permit a more direct comparison with theoretical models concerning the formation of radicals and complex molecules on sub-micrometre grains. Moreover, the timing of our SAAO observations allows some quantitative comparisons of the production rate of the C₂ and CN radicals and of the dust grains. Our analysis suggests that the model of evaporation from the CHON particles may be limited by the estimated mass ejection rate of the sub-micrometre grains (<10³ g s⁻¹), a potential difficulty noted previously². The mechanism of direct ejection of CN and C₂ radicals from the surface of the nucleus¹⁴, on the other hand, has to be evaluated according to the temporal variations of the jet features under the influence of solar radiation pressure and nucleus rotation. Clarification must await detailed model study. The presence of a strong anisotropy in the [O I] distribution, directly related to the outgassing pattern of the water vapour, is quite clear in our filtered image specially designed for imaging [O I] emission. A comparison of the observed asymmetry with theoretical models²⁴⁻²⁷ will be essential in the understanding of the surface sublimation of comet Halley at the time of the Giotto encounter.

We thank all SAAO staff, and Dr K. Schmidt, E. Krieg and E. Mikusch of DFVLR. We also thank Mr H. Lauche at MP Ae

for providing the [O I] filter set and technical advice. The campaign in South Africa was financed by the Observatory of Padova on behalf of CNR.

Received 16 November 1987; accepted 22 January 1988.

1. A'Hearn, M. F. *et al.* *Nature* **324**, 649-651 (1986).
2. A'Hearn, M. F. *et al.* *Eur. Space Ag. spec. Publ.* **250**(1), 483-486 (1986).
3. Cosmovici, C. B., Mack, P., Schwarz, G., Craubner, A. & Ip, W. I. *Eur. Space Ag. spec. Publ.* **278**, 195-207 (1987).
4. Cosmovici, C. B., Mack, P., Craubner, H. & Schwarz, G. *Eur. Space Ag. spec. Publ.* **250**(2), 151-155 (1986).
5. Cosmovici, C. B. *et al.* *Eur. Space Ag. spec. Publ.* **250**(2), 375-379 (1986).
6. Ip, W.-H., Cosmovici, C. B., Mack, P., Craubner, A. & Schwarz, G. *Eur. Space Ag. spec. Publ.* **278**, 271-275, (1987).
7. Grün, E. *et al.* *Nature* **321**, 144-147 (1987).
8. Sagdeev, R. Z. *et al.* *Nature* **321**, 262-266 (1986).
9. Keller, H. U. *et al.* *Nature* **321**, 320-326 (1986).
10. Sekanina, Z. & Larson, S. M. *Nature* **321**, 357-361 (1986).
11. Wilhelm, K. *Nature* **327**, 27-31 (1987).
12. Sekanina, Z. *Adv. Space Res.* **5**, 307-316 (1986).
13. Kissel, J. *et al.* *Nature* **321**, 336-337 (1986).
14. Combi, M. R. *Icarus* **71**, 178-191 (1987).
15. McDonnell, J. A. M. *et al.* *Nature* **321**, 338-341 (1986).
16. McDonnell, J. A. M. *et al.* *Eur. Space Ag. spec. Publ.* **250**(2), 25-38 (1986).
17. Ip, W.-H. *Adv. Space Res.* **5**, 233-245 (1986).
18. Festou, M. C. *et al.* *Nature* **321**, 361-363 (1986).
19. Krasnopolsky, V. A. *et al.* *Eur. Space Ag. spec. Publ.* **250**(1), 459-464 (1986).
20. Roesler, F. L. *et al.* *Eur. Space Ag. spec. Publ.* **250**(1), 413-415 (1986).
21. Spinrad, H., McCarthy, P. J. & Strauss, M. A. *Eur. Space Ag. spec. Publ.* **250**(1), 437-438 (1986).
22. Krankowsky, D. *et al.* *Nature* **321**, 326-329 (1986).
23. Gringauz, K. I. *et al.* *Nature* **321**, 282-285 (1986).
24. Kitamura, Y. *Icarus* **66**, 241-257 (1986).
25. Kömle, N. & Ip, W.-H. *Eur. Space Ag. spec. Publ.* **250**(1), 523-527 (1986).
26. Kömle, N. & Ip, W.-H. *Eur. Space Ag. spec. Publ.* **278**, 247-254 (1987).
27. Kitamura, Y. *Icarus* (in the press).

Comparison of SAGE II solar extinction data with airborne measurements of atmospheric backscattering in the troposphere and lower stratosphere

J. M. Vaughan*, D. W. Brown*, P. H. Davies*,
C. Nash*, G. Kent† & M. P. McCormick‡

* Royal Signals and Radar Establishment, St Andrews Road,
Great Malvern, Worcestershire WR14 3PS, UK

† Science and Technology Corporation, 101 Research Drive,
Hampton, Virginia 23666, USA

‡ NASA Langley Research Center, Hampton, Virginia 23665, USA

Since October 1984 the SAGE II limb sounding satellite has been measuring the extinction $\sigma(\lambda)$ of the Earth's atmosphere in seven wavebands. The extinction at the longest wavelength, $\sigma(1.02 \mu\text{m})$, is largely due to particulate material (aerosols) in the atmosphere. It is of considerable interest to relate these extinction values to the strength of atmospheric backscattering at this and longer wavelengths. In particular, atmospheric backscattering coefficients $\beta(\pi, \lambda)$ in the 9-11 μm range are important because projected performance of a satellite-borne CO₂ lidar for global wind measurement depends critically on their magnitude (the π signifies backscatter). In April 1986, during passage of the SAGE II satellite, the first simultaneous measurements of $\beta(\pi, 10.6 \mu\text{m})$ were made with an airborne CO₂ lidar. Individual ratios of $\beta(\pi, 10.6 \mu\text{m})$ and SAGE II extinction $\sigma(1.02 \mu\text{m})$ are in reasonable accord with previously calculated values, but the trend with height shows a distinctly nonlinear relation, which is probably attributable to a steadily changing size distribution of aerosols.

The NASA Stratospheric Aerosol and Gas Experiment (SAGE II)¹ on the Earth Radiation Budget Satellite (ERBS) is in a 610 km altitude circular orbit with a 57° inclination yielding global coverage ranging between 80° S and 80° N latitude every few weeks. Extinction measurements of solar radiation are made twice in each orbit during spacecraft sunrise and sunset (units of extinction, m⁻¹). The satellite orbital period is ~100 min so there are approximately 30 sampling opportunities per day.

Table 1 Comparison of local-atmospheric-backscatter coefficient (from the aircraft) and long-path extinction (from the satellite)

Altitude (km)	$\beta(\pi, 10.6 \mu\text{m})$ (10 ⁻¹⁰ m ⁻¹ sr ⁻¹)	$\sigma(1.02 \mu\text{m})$ (10 ⁻⁷ m ⁻¹)	$\beta(\pi, 10.6 \mu\text{m})$ $\sigma(1.02 \mu\text{m})$ (10 ⁻⁴ sr ⁻¹)
6.5*	530.8	36.93	14.37
7.5	11.83	19.15	6.18
8.5	4.558	13.54	3.37
9.5†	2.198	11.39	1.93
10.5*	1.551	9.318	1.66
11.5*	1.132	8.563	1.32
12.5*	1.066	8.503	1.25

Data recorded at sunset on 21 April 1986. The $\beta(\pi, 10.6 \mu\text{m})$ values are averages over 1 km height intervals for comparison with the satellite extinction values of $\sigma(1.02 \mu\text{m})$ which are deconvolved with a spatial-height resolution of 1 km.

* Cloud-contaminated value.

† Values in the stratosphere.

During a sunrise or sunset event, measurements of solar radiation are taken from a tangent height of 150 km (where there is no atmospheric attenuation) to the Earth's surface (or until the Sun is obscured by clouds). The solar radiance data are mathematically inverted into values of atmospheric extinction versus tangent altitude (at 1 km height intervals) and with approximately 1 km vertical resolution; the corresponding horizontal viewing paths are approximately 250 km long. These total extinction profiles are then separated into aerosol-extinction and gaseous-constituent-concentration profiles.

On three successive days, 19-21 April 1986, SAGE II measurements were made at sunset near the United Kingdom. The RSRE True Airspeed (LATAS) lidar was flown as close as possible to the atmospheric paths observed by the satellite. The LATAS equipment was installed aboard an HS 125 aircraft (maximum ceiling 13 km above sea level) operated from the Royal Aircraft Establishment (Bedford). The lidar, which incorporates a continuous-wave CO₂ laser at 10.6 μm of ~4 W output, has been operated for over five years with good reliability; extensive measurements of true airspeed, windshear, microbursts, turbulence and pressure error have been made and reported with a description of the complete equipment². The continuous-wave

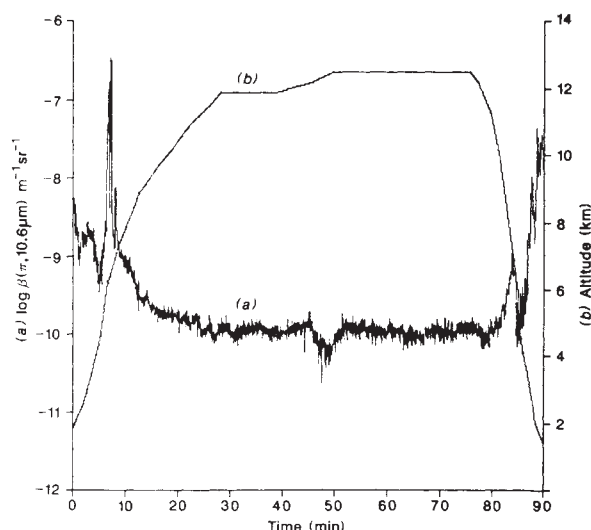


Fig. 1 Temporal record of the flight made at sunset on 21 April 1986. *a*, Backscatter coefficient $\beta(\pi, 10.6 \mu\text{m})$ referred to the left-hand ordinate scale. *b*, Aircraft altitude referred to the right-hand ordinate scale.

laser beam is transmitted from a 150 mm diameter telescope lens and for backscatter measurements is usually set to focus at ~ 100 m range directly ahead of the aircraft. This gives a depth of focus of about ± 15 m which, together with the beam diameter of ~ 10 mm at focus, defines the instantaneous, pencil-like, probe volume from which the bulk of the aerosol scattering emerges. Doppler-shifted radiation backscattered from aerosols re-enters the telescope and is subject to heterodyne detection and spectral analysis. Backscatter coefficients (units, $\text{m}^{-1} \text{sr}^{-1}$) are derived from the measured spectra from the calibration factors for the equipment. Spectra are usually accumulated over periods between 0.2 and 0.8 s and taken every few seconds. Thus from the speed of the aircraft the individual backscatter coefficients represent measurements over distances up to ~ 160 m depth.

Figures 1 and 2 are records of the flight made at sunset on 21 April 1986. In Fig. 3 backscatter values for each of the three flights are shown with extinction values due to SAGE II. The plotted values of $\beta(\pi, 10.6 \mu\text{m})$ have been obtained by taking arithmetic average of all the measurements (ascent, transit and descent) that fall within a given height interval. Salient features of each event are briefly outlined and discussed.

For the first event (flight 039; 19 April 1986) the tangent point at 10 km altitude for satellite observation was at 54.0°N , 5.2°E . This is in a busy airway and made any lengthy transit of the aircraft and lidar along the atmospheric path observed by the satellite impossible. Instead a restricted pattern was flown in a limited region of ~ 60 km diameter at the westerly end of the satellite viewing line. No obvious cloud was evident above ~ 2 km. However, during both ascent and descent two marked

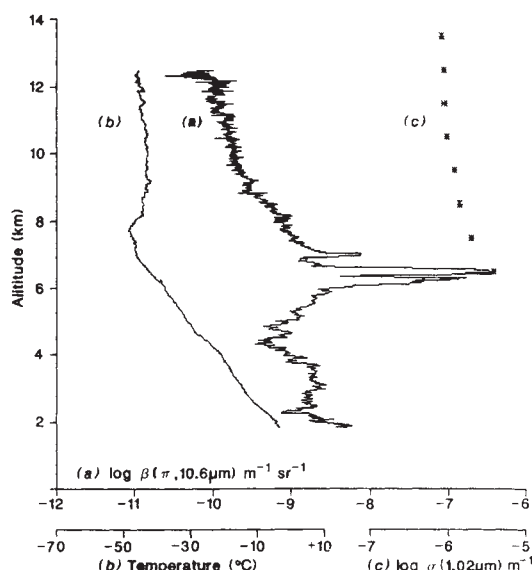


Fig. 2 Ascent portion of the flight made at sunset on 21 April 1986. *a*, Backscatter coefficient $\beta(\pi, 10.6 \mu\text{m})$. *b*, Static air temperature recorded at the aircraft. The points labelled *c*, are the contemporary values of atmospheric extinction $\sigma(1.02 \mu\text{m})$ for the SAGE II evaluation at 1 km height intervals.

layers of very strong scattering ($\sim 3 \times 10^{-9} \text{m}^{-1} \text{sr}^{-1}$) occurred at 7 and 9 km height). Colour photographs of the horizon made from the aircraft at this time show very distinct uniform dark layers (possibly due to incipient cirrus) extending well over 100 km north and south of the sunset. It is tempting to associate these with the backscatter layers. Because of these layers, averaging for comparison with SAGE data is made difficult, and was, in fact, done for height intervals of 2 km (Fig. 3*a*). The data points (crosses) for the stratosphere (above 9.5 km) are likely to be reliable.

For the second event (flight 040; 20 April 1986) the tangent point at 10 km altitude for satellite observation was at 51.2°N , 1.1°E . This too is in a busy airway and racetracks of 50 km length were flown ~ 50 km north of the satellite observation path. On this occasion patchy cloud extended up to the low tropopause at ~ 9 km. The averages were taken over 1 km height intervals; the data points in Fig. 3*b* for the stratosphere are likely to be reliable.

For the third event (flight 041; 21 April 1986) the tangent point at 10 km altitude for satellite observation was 48.2°N , 3.0°W . It was possible to fly at 12.5 km altitude for ~ 150 km near to the westerly half of the satellite observation path. Backscatter levels during the descent were very similar to those during the ascent (Fig. 2) down to 7 km altitude, but the strong peak below that was not so marked. This strong peak is almost certainly influenced by cloud. Averages were taken over 1 km height intervals and of the seven points shown in Fig. 3*c* the

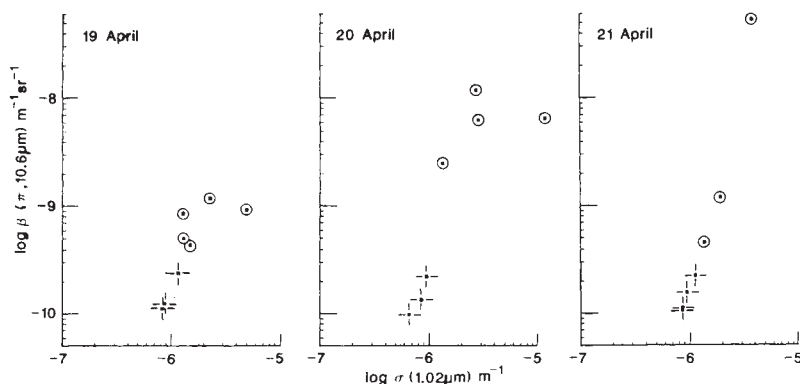


Fig. 3 Comparison of backscatter coefficient $\beta(\pi, 10.6 \mu\text{m})$ measured by coherent laser radar and the contemporary values of extinction $\sigma(1.02 \mu\text{m})$ due to SAGE II on three successive days (19–21 April 1986) near the United Kingdom, plotted at 1 km height intervals. The data points shown by a cross are in the stratosphere; the others (circled) are at the (low) tropopause or below.

lowest six for the heights 7.5 to 12.5 km should be reliable. The points extend through the tropopause and show a remarkably uniform trend. It is in fact possible to put a reasonably straight line between them but this is far from being at the 45° to the long axes required for a linear relation.

Comparison of the results in Fig. 3 shows that the three sets of data points in the stratosphere (marked by a cross) are in good agreement as to their position and change with altitude; but only for 21 April are the data at and below the tropopause down to 7.5 km free of either cloud or layers. Attention is accordingly focused on this set for which the absolute values and ratios of β (π , 10.6 μm) and σ (1.02 μm) are shown in Table 1. These experimental ratios of backscatter and extinction, ranging between $1.25 \times 10^{-4} \text{ sr}^{-1}$ at 12.5 km and $6.18 \times 10^{-4} \text{ sr}^{-1}$ at 7.5 km, are individually in fair accord with values calculated in other studies³. But the most obvious interpretation of the trend of the ratios with altitude is of a steadily changing aerosol size distribution that is reducing the backscatter with altitude much faster than the extinction. This could be due to a reduction of larger ($\geq 1 \mu\text{m}$) particles which contribute heavily to the 10.6 μm backscatter, with little change in the numbers of sub-micrometre particles which dominate the extinction. These findings emphasize the difficulty of assigning likely backscatter coefficients to observed extinction values without qualification in terms of altitude, time and geographic location. The results do show that there exists, on occasion, a reproducible relation between measurements made at the two wavelengths and provide an incentive for future comparison under different temporal and geographical conditions. As a final point it is worth noting that in contrast to Fig. 2 (and comparable records for the other flights) most previous airborne measurements⁴ have shown an increase of backscatter for the first few kilometres into the stratosphere. It will be interesting to see if this result is duplicated in future studies.

We thank R. Callan of RSRE for extensive assistance with the data recovery. G.S.K. has been supported by NASA.

Received 12 October 1987; accepted 27 January 1988.

1. Mauldin, L. E., Saun, N. H., McCormick, M. P., Guy, J. H. & Vaughan, W. R. *Opt. Engng* **24**, 307–312 (1985).
2. Woodfield, A. A. & Vaghan, J. M. *Int. J. Aviation Safety* **1**, 207–224 (1983).
3. Kent, G. S., Wang, P. H., Farrukh, U. O., Deepak, A. & Patterson, E. M. *Final Report on NASA Contract NAS 8-35594* (1985).
4. Vaughan, J. M. *et al.* *RSRE Rep.* no. 87002 (1987).

Negative octahedral snowflakes or Tyndall figures in tetrahydrofuran clathrate hydrate

Graham E. McLaurin & Edward Whalley

Division of Chemistry, National Research Council of Canada, Ottawa, Ontario, K1A 0R6 Canada

When a piece of ice Ih whose grains are not too small is irradiated by infrared or radio-frequency light, internal melt figures are often formed. They are often called Tyndall figures or Tyndall flowers after John Tyndall, who was the first to understand them¹. No other substance has been reported to form internal melt figures, although any substance that melts to a denser liquid should form them at the proper conditions. The shape of the figures will, of course, depend on the symmetry of the crystal. Tetrahydrofuran clathrate hydrate², which has the nominal composition $\text{C}_4\text{H}_8\text{O} \cdot 17\text{H}_2\text{O}$, melts to a denser homogeneous liquid at 4.1 °C with a fractional change of volume of -0.0147 ± 0.0003 ³. Internal melt figures should therefore form inside the clathrate if it is heated internally. From the symmetry and the structure of the crystal, the figures should be octahedra². This letter describes this

second type of internal melt figure, which was discovered 130 years after the first, and suggests that melt figures should occur in many other crystals.

Tyndall figures are formed when the interior of a crystal of ice is heated above its melting point by infrared radiation¹ or dielectric heating⁴. Because the melt is denser than the crystal, the crystal can melt in the interior by the conduction of heat to suitable nucleation sites. The resulting melt figures are flat hexagonal plates if the irradiation is not too strong, and are highly dendritic if the radiation is strong, for much the same reason that snow flakes are often highly dendritic—namely that the diffusion of heat and of matter follow similar laws. Because the melt is denser than ice, a small vapour bubble accompanies the liquid in the melt figure. Melt figures are often seen by those of us who climb glaciated mountains. The bare ice in the lower reaches of the glaciers is sometimes smooth enough on the surface—particularly in the walls of crevasses and melt-water tunnels—that the interior can be easily seen. The ice crystals are often quite large because the ice has been long annealing, and internal melt figures can be seen with the naked eye, or, better, with a lens of moderate power. It was by this method that they were first discovered¹.

In our experiments, distilled water was degassed by boiling, and reagent-grade 99.5% tetrahydrofuran was distilled to remove the quinol that was used as a stabilizer. They were mixed in the mole ratio 17:1. About 100 cm³ of the solution was transferred to a cylindrical glass freezing vessel that was about 20 cm long and 4 cm in diameter and that was joined to a narrow tube at the lower end and to a stopcock at the upper. The vessel was evacuated by opening the stopcock briefly several times at intervals of a few minutes and shaking vigorously, in order to remove most of the air without changing the composition significantly. The vessel and contents were cooled to 0 °C by immersing them in an ice bath, wiped dry, and quickly encased in Styrofoam insulation and placed vertically in a domestic refrigerator. Freezing of the solution was started by holding a small piece of solid carbon dioxide to the end of the glass tip of the vessel, and the sample was left in the refrigerator at -1 ± 1 °C to freeze slowly over a weekend in order to produce large single crystals of the clathrate. The final sample was completely frozen and almost clear.

The insulation was removed, and the vessel was mounted on a stand inside the refrigerator that allowed it to be rotated and moved vertically. The front of the refrigerator was then sealed with a 100- μm -thick polyethylene sheet. A microscope having a focal length of 6 cm was mounted directly in front of the refrigerator, and its objective lens was sealed into a hole that was cut in the polyethylene sheet and opposite the centre of the vessel. Arm holes were cut in the plastic sheet to allow the vessel to be moved relative to the fixed microscope, and were taped closed when not in use. The flexible plastic sheet permitted the focusing of the microscope on any part of the sample.

In order to generate the melt figures, a 250 W 'Heat Ray' infrared lamp that was placed close to the plastic sheet was switched on for 20–30 s. The sample was then illuminated with a microscope lamp pointed at a white paper that was behind the sample, and was examined with an American Optical AO 580 microscope.

The first irradiation produced many octahedral melt figures throughout the sample that were filled with liquid except for a small bubble at the upper end of each. The contrast between the liquid and the ice was not very great in arbitrary illumination because their densities differ by only about 1.5%³, but the octahedra were clearly seen when the illumination was adjusted properly. Photographs of the figures in four different crystals are reproduced in Fig. 1. They were taken at a total magnification of 19 with a Polaroid Land camera film pack mounted on the microscope, and a scale is drawn in Fig. 1c.

All the melt figures of the clathrate that were seen in this work, several hundred of them, as is illustrated in Fig. 1b, were

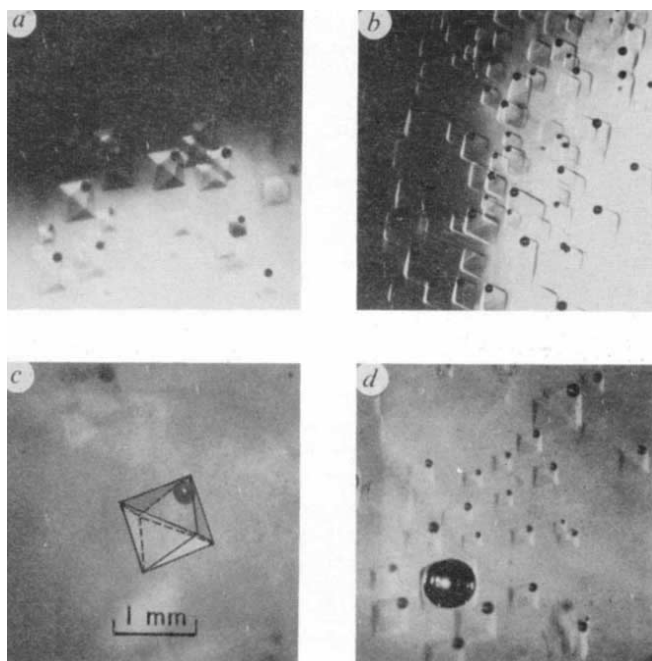


Fig. 1 Photographs of internal melt figures taken with a Polaroid camera. The scale of all the frames is drawn on frame *c*. All the melt figures are octahedra, and the larger octahedron in frame *c* is outlined in ink to help it to be easily seen. The small circular bubbles are the vapour spaces and the large bubble in frame *d* is on the surface of the clathrate and in the melt.

octahedra. This is as expected because tetrahydrofuran clathrate hydrate forms a cubic crystal and the octahedral planes are densely packed and so have a low energy². Typically, the edges of the octahedra were about 0.1–1 mm long. The bubbles seemed to be very nearly spherical, which is not surprising because of the high surface tension and the small size. The diameter of typical bubbles made in this work was about 20–200 μm .

The bubbles occur in the melt figures because the liquid is denser than the clathrate and so leaves a space filled with water vapour and tetrahydrofuran vapour. The ratio of the volume of the bubble, assuming it is spherical, to the volume of the melt figure is $\pi(d/l)^3$, where d is the diameter of the bubble and l the length of the diagonals of the octahedron. The dimensions of several octahedra and their bubbles were measured on several-fold enlargements of Fig. 1 and the average ratio of the volume of the bubble to the volume of the octahedron was 0.018 ± 0.004 . The melt is, therefore, denser than the clathrate by this fraction, which agrees with Gough and Davidson's³ value of -0.01470 ± 0.00034 .

Several pairs of octahedra grew together, as is seen in Fig. 1*a*. When they did so, the bubble in the lower figure rose to join the bubble in the upper.

All crystals that melt with a decrease of volume should form internal melt figures if heat can be dumped directly into the interior without conducting it from the outside surfaces. They should occur in some other structure II clathrate hydrates such as those of 2,5-dihydrofuran⁵, propylene oxide⁶, chloroform⁷, and others. They should also occur in the many other crystals whose melt is denser than the crystal, such as gallium, germanium, silicon, indium antimonide, sphalerite, and many others. It should be particularly easy to make the melt figures in gallium because it melts at 29.8 °C and the liquid is about 3% denser than the crystal. Heat could be dumped into the interior by passing an electric current, and the figures might be detected by X-ray shadowing.

The melting presumably originates at major imperfections in the crystals, and the figures may provide a way of studying these imperfections, and some preliminary experiments have detected

imperfections in ice that were generated by fast neutrons and γ -rays⁸. The melt figures can, of course, be used for orienting crystals, as was suggested over 50 years ago for the figures in hexagonal ice⁹.

Knight and Knight¹⁰ have shown how to make 'negative' crystals in ice without any fluid in them. They drilled a narrow hole in a single crystal near the melting point, inserted a hypodermic needle, and evacuated through the needle. With enough patience and a good enough seal between the needle and an appropriate crystal, negative crystals of tetrahydrofuran clathrate hydrate and other volatile materials could be made well below the melting temperature.

Finally, all these crystals should undergo either regelation or material transport¹⁰, that is, at the melting temperature should freeze together when touched together, should pass a weighted wire, and should make good snowballs.

We thank Dr D. D. Klug for useful discussions.

Received 11 January; accepted 18 February 1987.

1. Tyndall, J. *Proc. R. Soc. A* **9**, 76–80 (1858).
2. Davidson, D. W. in *Water a Comprehensive Treatise* Vol. 2 (ed. Franks, F.) 115–234 (Plenum, New York, 1973).
3. Gough, S. R. & Davidson, D. W. *Can. J. Chem.* **49**, 2691–2699 (1971).
4. Kass, M. & Magun, S. Z. *Krist.* **116**, 354–370 (1961).
5. Hawkins, R. E. & Davidson, D. W. *J. phys. Chem.* **70**, 1889–1894 (1966).
6. Wickert, J. N., Tamplin, W. S. & Shank, R. L. *Chem. Engng Progr. Symp. Ser.* **2** **48**, 92–96 (1952).
7. Tamman, G. & Krige, G. R. J. *Z. anorg. allgem. Chem.* **146**, 179–195 (1925).
8. Müller-Krumthar, H. in *Physics of Ice* (eds Riel, N., Bullemer, B. & Engelhardt, H.) 132 (Plenum, New York, 1969).
9. Adams, J. M. & Lewis, W. *Rev. Sci. Instrum.* **5**, 400–402 (1934).
10. Knight, C. A. & Knight, N. C. *Science* **150**, 1819–1821 (1965).
11. Hobbs, P. V. *Ice Physics* Ch. 6. (Clarendon, Oxford, 1974).

Forces between surfaces bearing terminally anchored polymer chains in good solvents

Hillary J. Taunton, Chris Toprakcioglu,
Lewis J. Fetters* & Jacob Klein†‡

Cavendish Laboratory, Madingley Road, Cambridge, CB3 0HE, UK
* Exxon Research and Engineering Corporation, Annandale,
New Jersey 08801, USA

† Polymer Research Department, Weizmann Institute of Science,
Rehovot 76100, Israel

The adsorption of high polymers modifies the short-range forces between surfaces in liquid media and is widely used to stabilize or flocculate colloidal dispersions¹. Previous studies have shown that low surface-coverage by the polymer leads to long-range bridging attraction between the surfaces², whereas repulsion sets in at high adsorbance in good solvents^{3,4}. In addition, marked hysteresis has been observed for both adsorbed homopolymers and co-polymers on approach and rapid separation of the surfaces^{3–5}. Here we report the direct measurement of forces between terminally anchored, non-adsorbing polymer chains (polystyrene) in good solvents (toluene and xylene). The polymer layers are found to extend some $6R_g$ (where R_g is the unperturbed radius of gyration of the polymer) out from each surface, substantially further out than adsorbing homopolymers. Monotonically increasing repulsive forces are observed as the surfaces approach each other and there is no evidence of bridging attraction at low-surface coverage. In addition we find that the force law is reproducible and independent of compression/decompression rates, in marked contrast with adsorbed homopolymers.

‡ To whom correspondence should be addressed.

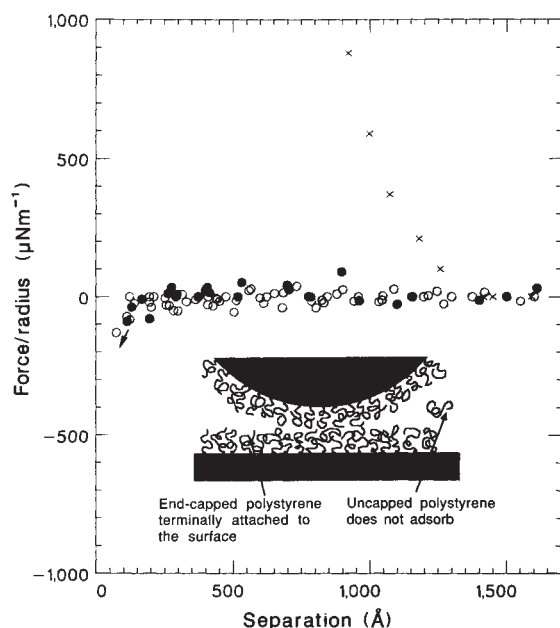


Fig. 1 Force-distance profiles for mica and polystyrene in xylene/toluene, measured using the force-balance technique. ●, Forces $F(D)$ between curved mica surfaces, radius of curvature R , in pure xylene (the result in toluene is identical). ○, Forces $F(D)$ after addition of unfunctionalized polystyrene and 12 h equilibration. The arrow is indicative of a jump into an adhesive contact, and occurs both before and after the addition of the polymer solution; crosses: 2 h after addition of end-capped polystyrene. The data are plotted as $F(D)/R$, which gives the corresponding interaction energy per unit area between flat plates distance D apart, in the Derjaguin approximation¹⁶.

The forces $F(D)$ between atomically smooth, crossed cylinders of mica, radius of curvature R (~ 1 cm) and distance D apart, were measured using the force-balance technique^{6,7}. The method also measures the refractive index of the medium separating the surfaces, and thus permits an estimate of the amount of adsorbed species.

The polymer/solvent systems used were polystyrene in toluene and polystyrene in xylene, both being good solvents for the polymer. The molecular characteristics of the polymer are given in Table 1. The mica-mica interaction in polymer-free solvent was in all cases measured as a control before the addition of polymer, and is shown in Fig. 1. The absence of any long range interaction and the jump into contact from about 100 Å (attributed to water-bridging in the undried solvent) are similar to previous studies⁴. Linear polystyrene (PS) was then added to the solvent to a concentration of 10^{-4} g ml⁻¹ and allowed to equilibrate with the mica surface for 12 ± 1 h. Previous investigations⁴ have indicated that polystyrene does not adsorb on to mica from undried toluene. The present study confirms this in both toluene and xylene, as the intersurface forces following equilibration with polystyrene are essentially unchanged from the pure solvent case (Fig. 1). But on replacement of the PS solution with a solution of the same concentration of end-functionalized polystyrene, PS-X (Table 1), where X is the zwitterion group $-\text{N}^+(\text{CH}_3)_2(\text{CH}_2)_3\text{SO}_3^-$, adsorption of the PS-X on to the mica was seen within 1 h as evidenced by the appearance of a repulsion in place of the marked attraction observed both with pure solvent and with the unfunctionalized PS. As the only difference between the two polymer chains is the presence of the zwitterion group in the end-functionalized polystyrene PS-X, this shows unambiguously that the PS-X chains are anchored to each mica surface by the highly polar X moiety, as indicated in Fig. 1.

Figure 2 shows the force-distance profile following saturation adsorption at different rates of approach and separation of the

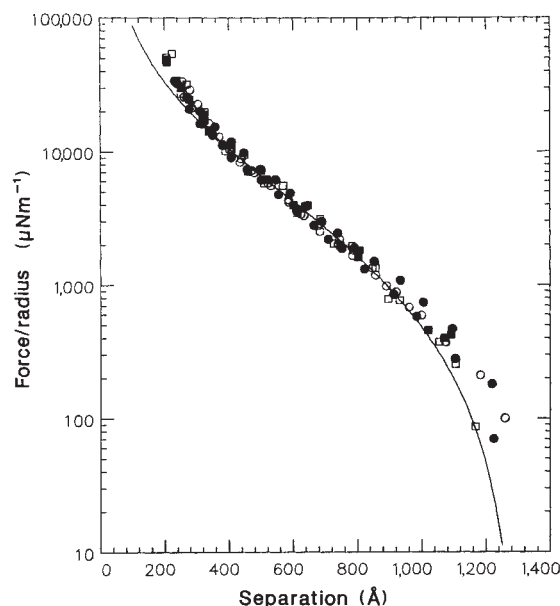


Fig. 2 Forces following saturation adsorption of PS-X. The squares and circles represent the results of two separate experiments. Solid symbols show the forces measured on compression of the adsorbed layers, and open symbols the corresponding forces on rapid decompression. Several compression/decompression cycles are shown. The solid line is the predicted energy $U(D)$ per unit area between flat plates a distance D apart, where $U(D) = \int_D^\infty F(D') dD'$ and $F(D)$ is given by equation 1 in the text. The value of L used in 650 Å, as measured from Fig. 1.

surfaces. The onset of repulsion is first detected at a separation of $1,250 \pm 70$ Å, indicating that the effective extension of the end-anchored polymer layers is ~ 650 Å (that is $\sim 6R_g$) from each surface. This highly extended configuration contrasts with measurements for adsorbed polymers in good solvents, where the extension is $\sim 3R_g$ from each surface²⁻⁴. Within the scatter, the profiles in Fig. 2 are independent of compression and decompression rates, again in contrast with adsorbed polymers, where marked hysteresis is a characteristic feature²⁻⁴. The large non-equilibrium relaxation effects observed with adsorbed chains in good solvents have been attributed to more monomers being forced on to the surface under compression⁴. In a system where the polymer does not adsorb but is only terminally attached, as in our study, no such effects would be expected.

Our results differ considerably from an earlier study of forces between surfaces in a solution of the block copolymer polystyrene/poly(vinyl-2-pyridine) (PS/PVP)^{5,8}, where the data were interpreted in terms of an adsorbing PVP block and a non-adsorbing PS block. Hysteresis was observed⁵ at moderate compressions ($>5,000 \mu\text{N m}^{-1}$) for a PS block with relative mass (M_r) of 1.85×10^5 (185 K), whereas repulsion was first noted at surface separations of $\sim 21 R_g$ (molecular radius of gyration of the PS block). In contrast, no hysteresis could be measured in our experiments, even under the most rapid

Table 1 Molecular characteristics of polymer

Polystyrene sample	M_r	M_r/M_N
PS	1.31×10^5	1.03
PS-X	1.41×10^5	1.02

The relative molecular masses were determined by gel-permeation chromatography. The unfunctionalized polystyrene PS, with microstructure $[\text{CH}_2\text{CH}(\text{C}_6\text{H}_5)]_n$, has a butyl and a phenyl group at its respective ends. The end-functionalized polystyrene, PS-X, was prepared using a propane sulfone¹⁵. M_r and M_N are the weight-averaged and number-averaged relative molecular masses respectively.

experimentally accessible compression and decompression rates (<10 min per cycle) and at very high repulsions ($50,000 \mu\text{N m}^{-1}$). Also, the range of repulsion in our case was very much smaller ($\sim 12R_g$ for a PS block of M_r 145 K). Identical force-distance profiles were measured in toluene and xylene, and the profiles remained stable for a period of up to 8 days. From refractive index measurements, we have determined the surface coverage of PS-X on mica to be $3 \pm 0.5 \text{ mg m}^{-2}$. We have not characterized the mica surface itself, but the interactions can be regarded as being due entirely to the end-grafted polymer chains at the separations of interest.

The equilibrium nature of our results permits detailed comparison with molecular theories of interaction between layers of end-anchored polymers⁹⁻¹². Scaling concepts have been used to describe the conformation of terminally adsorbed polymer layers in a good solvent¹⁰. This approach has been extended^{11,12} to obtain a form for the force per unit area, $F(D)$, between parallel flat plates a distance D apart and bearing a high coverage of end-grafted polymer chains

$$F(D) \approx \frac{kT}{s^3} \left[\left(\frac{2L}{D} \right)^{9/4} - \left(\frac{D}{2L} \right)^{3/4} \right] \quad (1)$$

for $D < 2L$, where k is Boltzmann's constant, T the absolute temperature, L the mean thickness of the end-grafted layer, and s the average distance between the anchor points on the surface. The force arises from two contributions: the first term in the bracket is the osmotic pressure due to the local increase in the polymer concentration between the plates as they are brought together, and the second term is the elastic restoring force.

This equation (corrected for the geometry of our experiment, see legend to Fig. 2) is plotted as the solid line in Fig. 2, shifted by a constant in the y axis, as the scaling argument does not give the numerical pre-factor. There is good agreement between theory and experiment over most of the range; the discrepancy at small separations may be due to the inapplicability of the scaling theory and volume exclusion effects at the high polymer concentrations in the intersurface gap.

The scaling model also gives an estimate of the layer thickness in a good solvent,

$$L = s \left[\frac{R_F}{s} \right]^{5/3} \quad (2)$$

where R_F is the end-to-end dimension of the polymer in a good solvent. We can evaluate the mean distance between anchor points, s , from the surface coverage estimated from our refractive index data, and R_F from intrinsic viscosity data^{13,14}: we find that equation 2 predicts L to be $740 \pm 50 \text{ \AA}$, which compares closely with our measured value of $650 \pm 50 \text{ \AA}$. Experiments varying both s and R_F (by changing M_r) are in progress.

This study shows directly that the interaction between layers of end-grafted, non-adsorbing polymer molecules in good solvent conditions is monotonically repulsive, non-hysteretic and considerably longer-ranged than for the corresponding adsorbing chains. The quantitative agreement of our results with the prediction of scaling models suggests that a scaling approach could be of considerable value in understanding the interactions between surfaces bearing terminally anchored polymer layers.

An SERC case studentship with Unilever (H.J.T.) and the partial support of the Israeli Academy Research Division (J.K.) are gratefully acknowledged. J.K. is the incumbent of the Herman Mark Chair of Polymer Physics. The authors thank T. M. Nicholson, and R. Yerushalmi-Rozen, and Prof. M. Tirrell and Prof. D. Tabor for comments on the manuscript.

Received 4 September 1987; accepted 5 February 1988.

6. Israelachvili, J. N. & Adams, G. E. *J. chem. Soc. Faraday I* **74**, 975-999 (1978).
7. Klein, J. *J. chem. Soc. Faraday I* **79**, 99-118 (1983).
8. Tirrell, M., Patel, S. & Hadzioannou, G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4725-4728 (1987).
9. Dolan, A. K. & Edwards, S. F. *Proc. R. Soc. A* **343**, 427-442 (1975).
10. Alexander, S. *J. Phys. Paris* **38**, 983-987 (1987).
11. De Gennes, P. G. *C. heb. Séanc. Sci.* **300**, 839-843 (1985).
12. De Gennes, P. G. *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca (1979)).
13. Roovers, J. E. & Bywater, S. *Macromolecules* **5**, 385-390 (1972).
14. Einaga, Y., Miyaki, Y. & Fujita, M. *J. Polym. Sci.* **17**, 2103-2109 (1979).
15. Davidson, N. S., Fetters, L. J., Funk, W. H., Graessley, W. W. & Hadjichristidis, N. *Macromolecules* (in the press).
16. Derjaguin, B. V. *Kolloid. Zh.* **69**, 155-164 (1934).

Long-term variation in seawater composition at the base of the thermocline

Francis J. Sansone, S. V. Smith*, James M. Price, Ted W. Walsh*, Thomas H. Daniel† & Christine C. Andrews

Department of Oceanography and Hawaii Institute of Geophysics and * Hawaii Institute of Marine Biology, University of Hawaii, 1000 Pope Road, Honolulu, Hawaii 96822, USA

† Natural Energy Laboratory of Hawaii, P.O. Box 1749, Kailua-Kona, Hawaii 96745, USA

Measurements of the chemical composition of seawater over a time scale of years at a fixed point in the ocean below the main thermocline have not been previously reported because of navigational limitations, ship and hydrowire instability, and, to a large extent, cost. The deep-water pipeline at the Natural Energy Laboratory of Hawaii at Keahole Point (Fig. 1a) provides a unique opportunity to address the question of time-varying water composition at a fixed site in the ocean below the mixed layer. In this study we examine the composition of sub-thermocline seawater through weekly sampling (surface-water data are also available, but are not discussed here). The mean deep-water composition obtained over about five years is representative of regional mean-water composition, and significant variation in water composition is observed on time scales exceeding one year. This unexpected long-term variation probably represents cyclic displacement of deep water masses in response to variation in the regional ocean circulation.

The most detailed long-term record of seawater chemistry known to the authors is the Panulirus Station 'S' data set, collected by the Bermuda Biological Station from south of Bermuda at monthly intervals from 1952 to the present¹. Although this record has proved valuable in many oceanographic studies, horizontal sampling locations are known only within kilometres, and depths are inferred from reversing thermometer data. In contrast, the Natural Energy Laboratory of Hawaii (NELH) has pumped sub-thermocline water virtually without interruption since mid-1982 through a pipe with an intake depth of 586 m, as part of an ocean-thermal energy conversion (OTEC) and aquaculture test facility (Fig. 1b). The intake of the pipe is located $25 \pm 5 \text{ m}$ above the rocky sea floor, and is 1,700 m offshore. The composition of this water has been measured roughly every week over that period, and additional shorter-term experiments have examined water composition over periods ranging from hours to 50 days. The present paper uses data from weekly samples collected between August 1982 and December 1987.

This paper addresses two primary questions. First, does the water composition obtained from the pumped seawater reflect the mean water composition in the region of the Hawaiian Islands? We will demonstrate that Keahole mean water composition closely matches the composition of open-ocean water in the Hawaiian region. This result will allow examination of the

1. Napper, D. H. *Polymeric Stabilisation of Colloidal Dispersions* (Academic, London, 1983).
2. Klein, J. & Luckham, P. F. *Nature* **308**, 836-837 (1984).
3. Klein, J. & Luckham, P. F. *Nature* **300**, 429-431 (1982).
4. Luckham, P. F. & Klein, J. *Macromolecules* **18**, 721-728 (1985).
5. Hadzioannou, G., Patel, S., Granick, S. & Tirrell, M. *J. Am. chem. Soc.* **108**, 2869-2876 (1986).

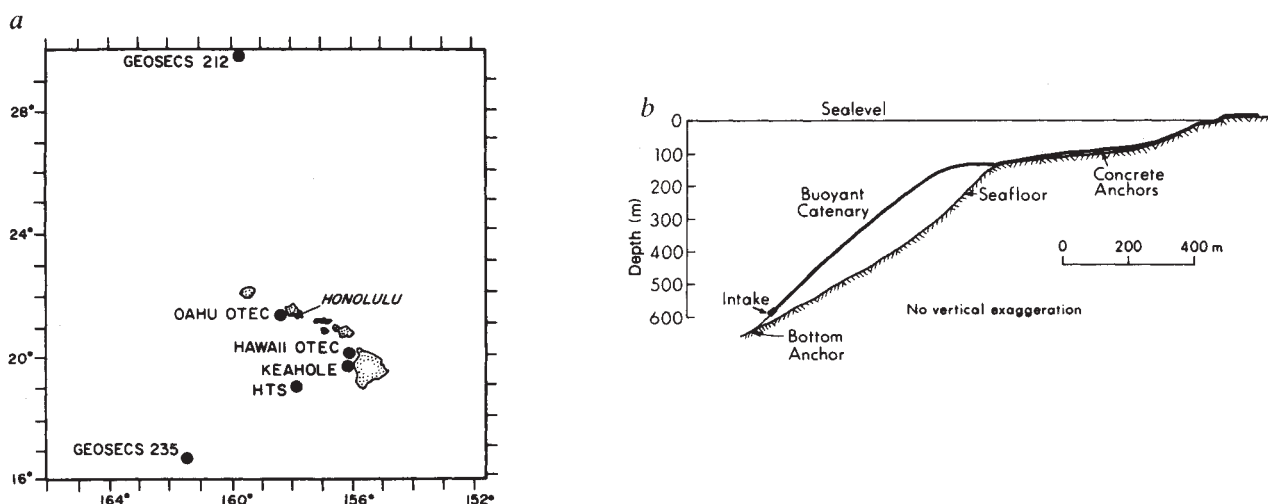


Fig. 1 a, Location map of the main Hawaiian Islands, showing the Keahole Point, Hawaii OTEC and Oahu OTEC sampling sites^{13,14}, Hawaii-Tahiti Shuttle (HTS)⁷, and GEOSECS stations 212 and 235¹². b, Cross-section, showing the configuration of the intake pipe at Keahole Point.

second question: what, if any, long-term trends occur in water composition at the base of the thermocline in the North Pacific Gyre?

The details of the analytical methods and spectral analyses which we have used here are described in a separate report on the short-term cyclic variations of the pumped seawater chemistry². This latter study observed significant, coherent periodicity in water composition ranging from ~6 h to 125 h. This short-term variability is interpreted as being due to vertical displacement of the water column by internal tidal oscillations, island-trapped waves and, possibly, remote meteorological forcing.

Table 1 summarizes the mean long-term data for temperature, salinity, and each of the inorganic plant nutrients commonly measured in hydrographic profiles. To filter high-frequency oscillations observed in the short-term study from the long-term trends, we have averaged the Keahole Point weekly data with an 11-point triangular running-mean filter. This filter is sufficient to remove signals with periods up to ~2.5 months—about the time scale of mesoscale eddies observed in Hawaiian waters^{3,4}.

Two main observations emerge from the inter-site comparison in Table 1. First, the mean salinity and nutrient concentrations at Keahole Point do not differ significantly from the means for the nearby sites; these sites range from being questionably

influenced by effects arising from the proximity of the Hawaiian islands⁵ (the two OTEC sites, within ~10 km of shore), to being well removed from any 'island mass' effect (Honolulu-to-Tahiti Shuttle and GEOSECS 235, 200–500 km from shore). GEOSECS station 212, 1,100 km north of Keahole Point, does not resemble the Keahole station in water composition, nor does it resemble the other tabulated sites. Because Keahole so closely resembles the other stations listed, it is likely that the discrepancy with respect to GEOSECS 212 represents a regional change between locations 1,500 km apart, rather than a local 'island mass' effect at Keahole Point.

The second important observation from these data concerns the variability in water composition observed at Keahole Point. The weekly data were filtered as described above before the tabulated standard deviations were calculated. Therefore, the variation reported in Table 1 cannot be explained by high-frequency oscillations in the water. We conclude that mean water composition at Keahole Point is representative of oceanic water composition near the base of the thermocline in the Hawaiian region, and that the observed variability is associated with some long-term variation in regional mean water composition at this specific site within the North Pacific Subtropical Gyre.

Figure 2a–d shows the filtered long-term records for salinity

Table 1 Comparison of the mean Keahole deep-water temperature, salinity and nutrient data with other data sources

	Keahole 586 m	Hawaii OTEC 453 m 679 m [586 m]	Oahu OTEC 456 m 679 m [586 m]	GEOSECS 235 598 m	GEOSECS 212 570 m	HTS 590 m
Temperature (°C)	9.4 ± 0.6	7.4 5.3 [6.1]	7.3 5.0 [6.0]	6.3	6.6	6.0
Salinity (‰)	34.30 ± 0.02	34.18 34.39 [34.30]	34.15 34.38 [34.28]	34.46	34.01	34.30
Nitrate (μM)	39.0 ± 0.8	33.5 41.9 [38.4]	30.9 39.1 [35.7]	41.3	29.5	37.0
Phosphate (μM)	2.95 ± 0.05	2.52 3.13 [2.88]	2.28 2.94 [2.66]	2.86	2.14	2.9
Silicate (μM)	74.4 ± 2.7	57.2 93.0 [78.3]	53.0 89.9 [74.5]	73.1	54.5	75.0

Keahole data are mean values ± 1 s.d. for weekly samples collected between 1982 and 1987. Note that the standard deviations are based on the filtered data, not the original data. Temperature data at Keahole Point are measured in the laboratory and are not representative of the ocean at the pumping depth because water warms in the pipe as a function of both flow rate (which is variable) and temperature of the mixed layer (which varies over an annual cycle). Short-term records from local hydrocasts and submersible dives indicate that the water temperature near the intake is ~6 °C, close to that of the offshore stations. Bracketed values listed under the OTEC site data represent linear interpolations of the data to 586 m. Locations and citations for the comparison sampling sites are given in Fig. 1a.

and each of the nutrients. The patterns which emerge indicate that there are long-term trends in water composition, and that the main trend is longer than annual. In particular, a decrease is observed in nutrient concentrations during 1983 and 1984, followed by a general increase since that time. In addition, we can compare the observed trend to the monthly mean sea level at Honolulu (K. Wyrki, personal communication; Fig. 2e), a significant variable which is analytically independent of the Keahole record of water composition. Note that the period of elevated sea level near the end of 1984 coincides with the minimum in the nutrient records.

In addition to the inter-annual variation, annual periodicity is apparent in all the records, especially in the concentration of silica and in sea level. It is least apparent in the salinity record, which has the smallest signal. The annual variation is manifest throughout the observation interval, except during the major decrease and recovery in the nutrient concentrations (1983 to early 1985).

Spectral analysis of the non-filtered weekly data from 3 August 1982 to 24 March 1987 yielded two major results. First, significant coherence at the 95% confidence level was observed among the nutrients and salinity, typically at periods of 1–2 months and at periods of 1 yr and longer (the record length was not long enough to distinguish the annual period from longer-period harmonics). Thus, the nutrients seem to be good water mass tracers at this depth. Second, sea level at Honolulu correlated with the nutrients and salinity at periods of 1 yr and longer. This was more prominently seen in the sea level/silica and sea level/nitrate coherences because of the larger signal-to-noise ratios of the silica and nitrate records. Sea level at Honolulu has been shown to be an indicator of the currents offshore⁶; we therefore suggest that long-term variation in the ocean circulation may be responsible for the observed annual and interannual changes in the water chemistry (see below).

Twice the standard deviation in Keahole composition divided by the vertical composition gradients near 600 m is a rough measure of the vertical displacement which could account for the observed variation in composition. To make this calculation we use data from Wyrki and Kilonsky⁷. It can be seen in Table 2 that vertical displacements of ~50–70 m would be enough to account for the variation we have observed in water composition. The inverse relationship between the maximum in sea level in late 1985 and the low nutrient concentrations at that time is consistent with elevation of the thermocline in response to low-frequency baroclinic Rossby waves⁸.

Alternatively, the water-mass displacements may be horizontal. There are insufficient data on spatial variation of water composition near the base of the thermocline in the Hawaiian region to resolve the actual spatial scale of water-mass displacement implied by these data. But dividing twice the standard deviation in Keahole composition by the horizontal composition gradient between GEOSECS stations 235 and 212 (about 1,500 km apart) provides an estimate of possible horizontal motions implied. From Table 2 it can be seen that a north-south shift of 130–450 km could account for the observed variations in composition.

We recognize the uncertainty in these calculations, including

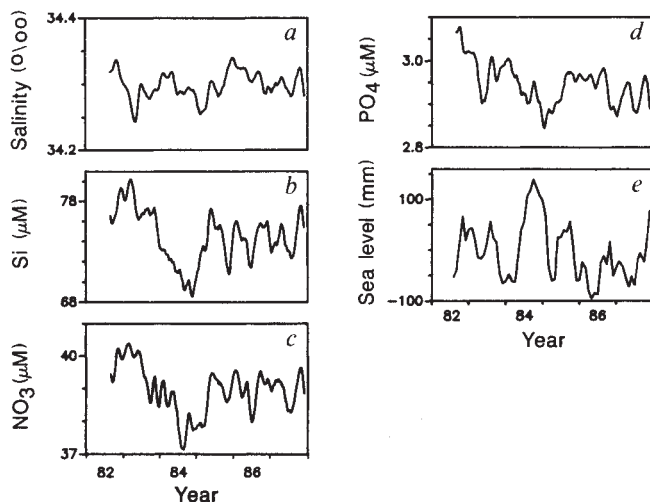


Fig. 2 a–d, Water composition at 586 m, Keahole Point, 1982–1987; these data have been smoothed with an 11-point triangular filter, as discussed in the text. e, Monthly mean sea level data at Honolulu, on the south coast of Oahu.

the possibility that the water-mass shift might be east-west rather than north-south. It is also likely that 'horizontal' gradients should actually be calculated along isopycnal surfaces, if sufficient data were available to resolve realistic trajectories of displacement. The calculations above use limited spatial data to evaluate the general magnitude of displacement in water-mass position needed to account for the temporal variation which is being documented at Keahole Point. It is significant, however, that the amplitude of the variation in the Keahole salinity record (Fig. 2a) is nearly identical to that reported by Seckel^{9,10} for interannual salinity variation in Hawaiian waters at this depth due to north-south movement of the southern boundary of the high-salinity (>35‰) North Pacific Central Water.

Although it is not possible to ascribe a detailed physical explanation to the observed evidence of variations in regional water-mass displacement¹⁰, one possibility is sufficiently obvious to bear mention. If, as seems possible from these data, the variation is cyclic on a time scale of a few years (Fig. 2), it may represent a response of the circulation of the North Pacific Gyre to phenomena on the same time scale as the El Niño/Southern Oscillation (that is, ~4 yr)¹¹. Hydrographic profiles with a spatial grid of tens of kilometres over an area a few hundred kilometres on a side would be needed to resolve the trajectory of water-mass displacement, and a continuation of the Keahole record will be important in resolving the time scale in question.

We thank R. Lukas, K. Wyrki and K. Chave for critical comments on the manuscript, B. Lee for water sampling and J. Bryan for graphics preparation. The United States Department of Energy, Ocean Energy Section and the State of Hawaii Department of Planning and Economic Development supported the water chemistry analyses and additional funding was provided by the United States National Science Foundation Chemical Oceanography Program.

Table 2 Inferred values for vertical and north-south horizontal displacement

	Keahole Point variability (2 s.d.)	Vertical gradient (change per m)	Vertical displacement (m)	Horizontal gradient (change per km)	Horizontal displacement (km)
Salinity (‰)	0.04	0.00078	51	0.00030	133
Nitrate (µM)	1.6	0.023	70	0.0079	202
Phosphate (µM)	0.10	0.0017	59	0.00048	208
Silicate (µM)	5.4	0.084	64	0.012	450

Displacement values were inferred by dividing twice the standard deviation (s.d.) of water composition at Keahole Point by either the vertical gradient from the HTS data⁷ or the horizontal gradient in composition between GEOSECS stations 235 and 212 (ref. 12).

Received 28 September 1987; accepted 19 February 1988.

- Jenkins, W. J. & Goldman, J. C. *J. mar. Res.* **43**, 465-491 (1985).
- Price, J. M., Sansone, F. J., Smith, S. V., Rosti, P. & Andrews, C. C. *J. mar. Res.* (submitted).
- Patzert, W. C. *Eddies in Hawaiian waters* Rept 69-8 (Hawaii Inst. Geophys., Honolulu, 1969).
- Lobel, P. S. & Robinson, A. R. *Deep Sea Res.* **33**, 483-500 (1986).
- Gilmartin, M. & Relevante, M. *J. exp. mar. Biol. Ecol.* **16**, 181-204 (1974).
- Price, J. M. *J. phys. Oceanogr.* **11**, 1375-1382 (1981).
- Wyrtki, K. & Kilonsky, B. *Transequatorial water structure during the Hawaii to Tahiti Shuttle Experiment* Rept 82-5 (Hawaii Inst. Geophys., Honolulu, 1982).
- Wyrtki, K. & Wenzel, J. *Nature* **309**, 538-540 (1984).
- Seckel, G. R. *Trans. Am. geophys. Union* **49**, 377-387 (1968).
- Seckel, G. R. in *HOE: The Hawaiian Ocean Experiment* (ed. Magaard, L., Pujalet, R. & Gaynor, V.), 41-60 (Hawaii Inst. Geophysics, Honolulu, 1985).
- Quinn, W. H., Zopf, D. O., Short, K. S. & Kuo Yang, R. T. W. *Fishery Res. Bull.* **76**, 663-678 (1978).
- Broecker, W. S., Spencer, D. W. & Craig, H. *GEOSecs Pacific expedition* Vol. 3 (National Science Foundation, Washington, DC, 1982).
- Noda, E. K., Bienfang, P. K., Kimmerer, W. J. & Walsh, T. W. *Environmental surveys during operation and following removal of the OTEC-1 system off Keahole Point, Hawaii, January-April 1981* (United States DOE/NBM-2016727, Washington, DC, 1981).
- Noda, E. K., Bienfang, P. K., Kimmerer, W. J. & Walsh, T. W. *OTEC environmental benchmark survey off Kahe Point, Oahu, May 1980-May 1981* (United States DOE/NBM-2016054, Washington, DC, 1981).

Dissolved gallium in the open ocean

Kristin J. Orians & Kenneth W. Bruland

Institute of Marine Sciences and Chemistry Board of Studies,
University of California at Santa Cruz, Santa Cruz,
California 95064, USA

Greater understanding of the distributions and biogeochemical cycling of trace metals in the oceans has recently been attained due largely to advances in analytical chemistry and in trace metal sampling. We present here the first oceanographically consistent distribution of dissolved gallium in the open ocean, with concentrations 1-2 orders of magnitude lower than those reported previously¹⁻³. The distribution of dissolved gallium is complex, indicating multiple sources and removal by scavenging throughout the water column.

Gallium, a trivalent metal, hydrolyses in seawater to form $\text{Ga}(\text{OH})_3^0$ and $\text{Ga}(\text{OH})_4^-$ (ref. 4). Below aluminium in the periodic table of the elements, gallium should have chemical behaviour similar to aluminium and to other readily hydrolysed elements such as iron, thorium and bismuth. Existing data on these hydrolysis elements is now sufficient to test theories on their particle reactivity and oceanic chemistry. The dissolved gallium data presented here is valuable for constraining these models.

The oceanic concentration of an element can be estimated based on its rate of supply to the oceans and its water column reactivity. By analogy with aluminium, the primary source of dissolved gallium to the open ocean may be atmospheric input of crustal dust. Assuming that gallium and aluminium have similar extents of dissolution from atmospheric particles and similar water column reactivities, the seawater concentrations of these two elements should be proportional to their concentration ratio in crustal dust. A dissolved gallium concentration of 0.2 pM is predicted using crustal abundances of gallium and aluminium⁵ and the average dissolved aluminium concentration in seawater reported by Orians and Bruland⁶. This estimated concentration is three orders of magnitude lower than the average value of 0.4 nmol kg^{-1} presented in a recent review⁴ based on the few dissolved gallium measurements reported in the late 1950s¹⁻³. New determinations of dissolved gallium in the oceans utilizing modern methodology were, therefore, warranted.

Seawater samples from the North Pacific central gyre (28°N , 155°W ; VERTEX IV, July 1983) and from the Gulf Stream in the northwest Atlantic (36°N , 73°W ; July 1987) were collected and filtered using well established procedures^{7,8} that allow collection of uncontaminated samples for trace metal analysis. The distribution of dissolved gallium with depth in the central North

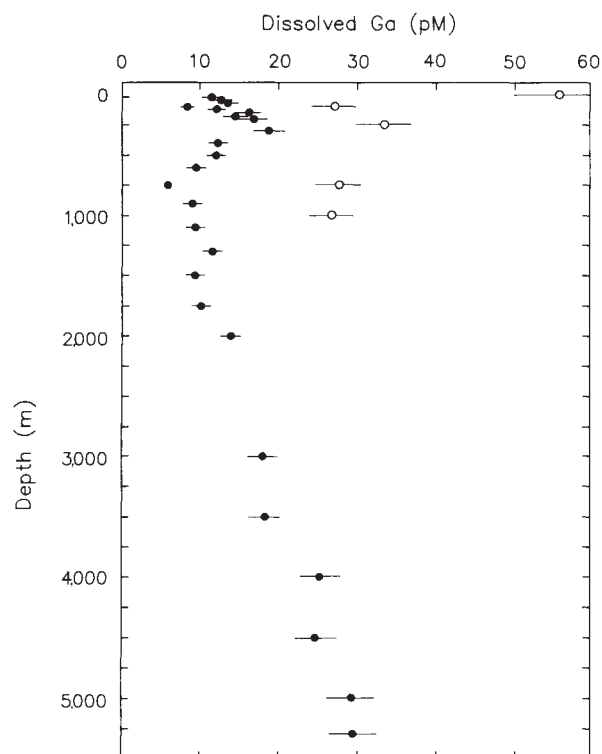


Fig. 1 Variation of dissolved gallium with depth at 28°N , 155°W (VERTEX IV, July 1983) in the central North Pacific ($\bullet$), and at 36°N , 73°W (Gulf Stream, July 1987) in the northwest Atlantic ($\circ$). Dissolved gallium in the Pacific samples was concentrated 80-fold at sea using a Chelex-100 ion-exchange technique⁸. An additional 10-fold evaporative concentration step was performed, resulting in a total concentration factor of ~ 800 . The Atlantic samples were concentrated 1,000-fold using an 8-hydroxyquinoline resin³⁰. Eluates were then analysed for dissolved gallium by graphite furnace atomic absorption spectrophotometry. The error bars represent the analytical precision (per cent difference for duplicate sample analysis) of 12%. The detection limit of these techniques is 1.2 pM (three times the standard deviation of the blanks, which were below the detection limit). Details of both techniques will be presented elsewhere.

Pacific gyre (Fig. 1) exhibits intermediate concentrations at the surface (12 pM), a subsurface maximum near 300 m (18 pM), a broad minimum from 500 m to 1,000 m (7-10 pM), and increasing concentrations with depth below 1,000 m to a maximum in the bottom waters (30 pM). The northwest Atlantic Gulf Stream samples (Fig. 1) have dissolved gallium concentrations roughly three times higher than those at comparable depths in the central North Pacific.

These values are 1-2 orders of magnitude lower than those in previous reports¹⁻³, but one to two orders of magnitude higher than predicted by direct comparison with aluminium. This indicates that gallium and aluminium have some important differences in their oceanographic behaviour. Compared to dissolved aluminium, dissolved gallium may have additional sources, may be dissolved more readily from atmospheric particles or may be less readily scavenged from the water column. Reduced reactivity would increase the oceanic residence time of dissolved gallium relative to dissolved aluminium and, therefore, lead to higher concentrations than otherwise predicted. Evaluation of the spatial distribution of dissolved gallium in the oceans will help in understanding the differences between dissolved aluminium and dissolved gallium, and whether additional sources or decreased reactivity lead to the higher than predicted dissolved gallium concentrations.

The vertical distribution of dissolved gallium in the central North Pacific (Fig. 1) suggests surface, subsurface, and deep

water sources, combined with particle scavenging throughout the water column. The source of dissolved gallium to the surface waters of the open ocean could be eolian or fluvial. Dissolved aluminium input has been found to be predominantly eolian^{6,9-12} but fluvial input has also been reported to be important in certain regions, such as in the northwest Atlantic intermediate and deep waters¹³ and in the Panama Basin¹². The high dissolved gallium concentration observed in the surface waters of the Gulf Stream could be due to Saharan dust input to the equatorial Atlantic, or fluvial input to the Gulf of Mexico, transported advectively by the Gulf Stream to the northwest Atlantic station. Additional data is clearly needed to elucidate the source of dissolved gallium to the surface ocean.

The subsurface maximum and broader mid-depth minimum could be advective features or the result of reversible exchange processes involving transient sinking particles. Dissolved gallium in the surface waters may be incorporated into some solid phase which undergoes shallow regeneration, such as biological soft tissue. After regeneration the dissolved gallium may then be rescavenged by particles, a process which occurs throughout the water column for particle reactive elements¹⁴. Subsurface concentration maxima have also been observed for dissolved bismuth¹⁵, plutonium 239, plutonium 240¹⁶ and anthropogenic lead¹⁷, but these occurred roughly 200 m deeper. The maxima for bismuth and plutonium were found to transgress density surfaces, implying a source more complex than simple lateral advection^{15,16}. Reversible exchange was proposed as a possible explanation. The source of the subsurface dissolved gallium maximum cannot be distinguished with the existing data.

There may be several sources of dissolved gallium to the deep waters (below 1,500 m): regeneration from particles in the water column; remineralization at the sediment surface; diffusion out of the sediments; and advective transport. The presence of a deep water source, as yet undefined, is useful for investigating the scavenging of trace elements in the deep ocean.

The deep water dissolved gallium distribution has features similar to those seen for copper^{8,18} and aluminium^{6,9}. The concave shapes evident when the concentrations of these dissolved metals are plotted against a conservative tracer, such as salinity or potential temperature, is evidence of deep water scavenging. Quantitatively, the extent of removal due to scavenging can be expressed as a deep water scavenging residence time. A simple, vertical advection/diffusion model¹⁹ has been used to estimate scavenging residence times for copper and aluminium of approximately 1,000 years¹⁸ and 150 years⁶, respectively. When applied to the dissolved gallium data from the central North Pacific gyre (Fig. 1), this model yields a deep water scavenging residence time of roughly 750 years. This demonstrates that dissolved gallium is less reactive to particle scavenging than dissolved aluminium in the deep waters, but more reactive than dissolved copper.

The three fold depletion of dissolved gallium in the intermediate waters (500–1,000 m) of the North Pacific, relative to the North Atlantic, is consistent with a scavenging reactivity intermediate between aluminium and copper. Elements exhibiting nutrient-type cycling are enriched in the older intermediate waters of the North Pacific compared to the North Atlantic; the five-fold enrichment of dissolved zinc in the North Pacific is the greatest nutrient-type inter-ocean fractionation yet observed for a trace metal²⁰. Particle reactive elements, on the other hand, tend to be depleted in the North Pacific; the 40-fold depletion of dissolved aluminium in the North Pacific is the largest scavenging-type inter-ocean fractionation found⁶. Elements between these two extremes are influenced, to various degrees, by nutrient-type cycling and particle-scavenging. Dissolved copper, for example, is enriched by a factor of 1.5–2.0 in the North Pacific relative to the North Atlantic^{18,20,21}, and iron is depleted three- to four-fold in the North Pacific²²⁻²⁴. Dissolved gallium appears to exhibit an inter-basin fractionation similar to that of dissolved iron.

The observed reactivity of dissolved gallium in the oceans can be used to test recent model predictions of trace metal reactivity. A model relating the hydrolysis side reaction coefficients of elements ($\log \alpha_{M,OH}$) with their polarizing power (as indicated by the ionic charge to radius ratio) describes hydrolysis-dominated elements as those with $\log \alpha_{M,OH}$ between two and twenty-five (beyond which condensation to oxy-anions occurs)^{25,26}. This model predicts that a hydrolysis-dominated element with a higher $\log \alpha_{M,OH}$ will be more particle reactive and, therefore, will have a shorter oceanic residence time. Dissolved gallium, according to this model, should have a shorter oceanic residence time than dissolved aluminium and iron, and should be very similar to thorium in reactivity. The data presented here does not support this hypothesis; additional factors must be involved.

A chemical equilibrium program²⁷, adapted for use on personal computers²⁸, using recently compiled hydrolysis constants²⁵, predicts that $Ga(OH)_4^-$ should be the dominant inorganic gallium solution species in seawater (95–99%). Dissolved aluminium, on the other hand, exists as a nearly equal mixture of $Al(OH)_3^0$ and $Al(OH)_4^-$, and dissolved iron and thorium exist predominantly as their neutral species, $Fe(OH)_3^0$ and $Th(OH)_4^0$. Since particle surfaces in seawater have a net negative charge²⁹, anionic metal species would need to lose a hydroxyl group in order to complex with a negatively charged donor atom on a particle surface. Exchange kinetics could thus become important and may decrease the surface adsorption reactivity of anionic species. The predominance of the anionic form of gallium could be one possible explanation for the longer oceanic residence time observed for dissolved gallium relative to aluminium, iron and thorium than that predicted simply by affinity for hydroxide ions.

In conclusion, dissolved gallium concentrations in the central North Pacific and upper 1,000 m of the northwest Atlantic range from 7 to 30 pM and 28 to 56 pM, respectively. Pacific ocean deep-water concentrations are higher than those in surface waters, and Atlantic concentrations (upper 1,000 m) are higher than those in the Pacific. The vertical distribution suggests complex controls on dissolved gallium in the oceans, with multiple sources, reversible exchange and scavenging processes all contributing to the observed oceanic chemistry. The estimated deep water scavenging residence time of 750 years and the observed inter-ocean fractionation demonstrate that dissolved gallium is less reactive in the oceans than one would anticipate based on its affinity for hydroxide ions. We propose that the reduced particle reactivity of dissolved gallium, relative to other hydrolysis elements, may be caused by the predominance of its anionic solution species, $Ga(OH)_4^-$ in seawater.

We especially thank Geoffrey Smith for handling the Chelex-100 sample processing at sea, Rob Franks for his help in collecting the North Pacific samples, Dr Greg Cutter for collecting the Atlantic samples, and Dr Margaret Delaney, Dr Gary Gill and John Donat for editorial help. This work was funded by the NSF, Marine Chemistry division.

Received 2 November 1987; accepted 18 February 1988.

1. Culklin, F. & Riley, J. P. *Nature* **181**, 179–180 (1958).
2. Burton, J. D., Culklin, F. & Riley, J. P. *Geochim. cosmochim. Acta* **16**, 151–80 (1959).
3. Ishibashi, M., Shigematsu, T., Nishikawa, Y. & Hiraki, K. *Nippon Kagaku Zasshi* **82**, 1141–1143 (1961).
4. Bruland, K. W. in *Chemical Oceanography*, Vol. 8 (eds Riley, J. P. & Chester, R.) 157–220 (Academic, London, 1983).
5. Taylor, S. R. *Geochim. cosmochim. Acta* **28**, 1273–1285 (1964).
6. Orians, K. J. & Bruland, K. W. *Nature* **316**, 427–429 (1985).
7. Bruland, K. W., Franks, R. P., Knauer, G. A. & Martin, J. H. *Analyt. chim. Acta* **105**, 233–245 (1979).
8. Bruland, K. W. *Earth planet. Sci. Lett.* **47**, 176–198 (1980).
9. Orians, K. J. & Bruland, K. W. *Earth planet. Sci. Lett.* **78**, 397–410 (1986).
10. Hydes, D. J. *Science* **205**, 1260–1262 (1979).
11. Hydes, D. J. *Geochim. cosmochim. Acta* **47**, 967–973 (1983).
12. Measures, C. I., Grant, B., Khadem, M., Lee, D. S. & Edmond, J. M. *Earth planet. Sci. Lett.* **71**, 1–12 (1984).
13. Measures, C. I., Edmond, J. M. & Jickells, T. D. *Geochim. cosmochim. Acta* **50**, 1423–1429 (1986).
14. Bacon, M. P. & Anderson, R. F. *J. geophys. Res.* **87**(C3), 2045–2056 (1982).

15. Lee, D. S., Edmond, J. M. & Bruland, K. W. *Earth planet. Sci. Lett.* **76**, 254-262 (1985/86).
16. Bowen, V. T., Noshkin, V. E., Livingston, H. D. & Volchok, H. L. *Earth planet. Sci. Lett.* **49**, 411-434 (1980).
17. Schaule, B. K. & Patterson, C. C. *Earth planet. Sci. Lett.* **54**, 97-116 (1981).
18. Boyle, E. A., Slater, F. R. & Edmond, J. M. *Earth planet. Sci. Lett.* **37**, 38-54 (1977).
19. Craig, H. *Earth planet. Sci. Lett.* **23**, 149-159 (1974).
20. Bruland, K. W. & Franks, R. P. in *Trace Metals in Seawater* (eds Wong, C. S., Boyle, E. A., Bruland, K. W., Burton, J. D. & Goldberg, E. D.) 395-414 (Plenum, New York, 1983).
21. Moore, R. M. *Earth planet. Sci. Lett.* **41**, 461-468 (1978).
22. Gordon, R. M., Martin, J. H. & Knauer, G. A. **299**, 611-612 (1982).
23. Symes, J. L. & Kester, D. R. *Mar. Chem.* **17**, 57-74 (1985).
24. Landing, W. M. & Bruland, K. W. *Geochim. cosmochim. Acta* **51**, 29-43 (1987).
25. Turner, D. R., Whitfield, M. & Dickson, A. G. *Geochim. cosmochim. Acta* **45**, 855-881 (1981).
26. Whitfield, M. & Turner, D. R. in *Aquatic Surface Chemistry* (ed. Stumm, W.) 457-494 (Wiley, New York, 1987).
27. Westall, J. C., Zachary, J. L. & Morel, F. M. M. *Technical Note* **18**, 1-91 (Department of Civil Engineering, MIT, Cambridge, 1976).
28. Cabaniss, S. E. *Environ. Sci. Technol.* **21**, 209-210 (1987).
29. Neihof, R. A. & Loeb, G. I. *J. mar. Res.* **32**, 5-12 (1974).
30. Landing, W. M., Haraldsson, C. & Paxeus, N. *Analyt. Chem.* **58**, 3031-3035 (1987).

A model for transition fields during geomagnetic reversals

I. Williams,* R. Weeks† & M. Fuller†

* Department of Plant and Earth Sciences, University of Wisconsin, River Falls, Wisconsin, 54022, USA

† Department of Geological Sciences, University of California, Santa Barbara, California 93106, USA

Palaeomagnetic studies of the reversal of the geomagnetic field have produced a number of palaeomagnetic records of intermediate or transition fields during reversals. These transition fields may provide an important clue as to the nature of the geodynamo. During the last reversal, the geomagnetic field was no longer dipolar¹, but the field appeared to be dominated by low-order harmonics²⁻⁴. Such transition-field geometries predict that all palaeomagnetic records will include high inclinations of either + or - 90°, which give the virtual geomagnetic pole paths passing through the site or its antipode. Subsequently, when transition records were found not showing these characteristics, the idea of the dominance of low-order zonals was questioned. Here we show that if the low-order zonals are combined with a drifting non-dipole field, similar to that seen at present, most of the features of reversal transitions can be simulated. We therefore suggest that low-order zonals may indeed dominate transition fields, but they are accompanied by some form of drifting non-dipole field.

In an earlier zonal harmonics model⁴, geomagnetic reversals were simulated by allowing the dipole to decay to zero and grow in the opposite sense, with the variation in dipole field energy controlled by a tangent function. The energy lost from the dipole field was fed into the axial quadrupole, octupole, and hexadecapole in the ratio of 20%, -30% and 50%, with the energy in the field outside the core being conserved.

The model demonstrated that individual records of a reversal would vary considerably, depending upon the latitude of the observation site. The axisymmetry of the model, however, predicted that records would be invariant with longitude. The zonal model was partially successful in simulating reversal records, but it was too simple to explain the variety of phenomena seen during reversals. The principal failing was that it could not account for records that did not include inclinations of + or -90°. In presentations of records in terms of virtual geomagnetic pole (VGP) paths, the zonal model predicts that the pole will always pass immediately under the site, or under its antipode (Fig. 1a). But records were observed without the necessary high inclination so that the VGP path was close to neither the site nor its antipode. An additional complication was that such paths were frequently confined to a small band of longitude but displaced from the site, or the antipodal meridian. This also suggested that a drifting non-dipole field could not alone be

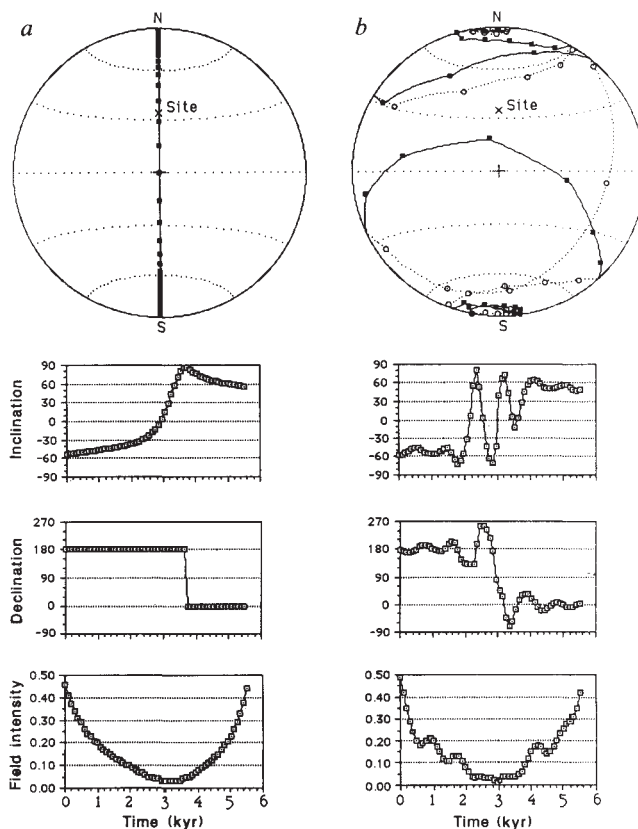


Fig. 1 Synthetic palaeomagnetic records of a reverse to normal transition of the geomagnetic field observed at 35° N. The upper-most row of plots are virtual geomagnetic pole (VGP) paths viewed from above the Equator where it intersects the site meridian. The lower profiles give the succession of inclination, declination and field intensity values as a function of time during the reversal. *a*, Zonal model. *b*, Drifting non-dipole field model.

responsible for the observed paths; such a field geometry defines a path which sweeps around the surface of the Earth (Fig. 1b).

The present extension of the zonal model involves the addition of a westward drifting non-dipole field. The 1980 non-dipole field was used and the drift rate was chosen to be similar to the observed 0.2°-per-year rotation. The intensity of the non-dipole field was comparable to the present drifting non-dipole field, which gave directional and intensity changes comparable with those seen in the records from Crete⁵ and Japan⁶.

Figure 2 illustrates the results for observation sites on a single line of latitude in the Northern Hemisphere, in the form of VGP paths and declination, inclination and intensity plots. Note that, as in the case of the last reversal, there are simple VGP paths confined to the site meridian, which would be interpreted as evidence of the dominance of zonals, but there are also others revealing the importance of non-zonals. The paths are controlled by the phase relationship of the variation of strength of the zonals and the drifting non-dipole field. Hence we see a wide variety in the paths along a line of latitude as a function of longitude.

The palaeomagnetic reversal data do not yet permit a useful inversion in terms of spherical harmonic content. It is clear, however, that the synthetic reversal paths include a number of features seen in the observed records. The preference for near-sidedness is due to the control of the path by the zonals even in the absence of a dipole, as recognized earlier². The drifting non-dipole field can displace the VGP path in longitude away from the site meridian and, providing that the main directional change takes place when the non-dipole field is not changing rapidly, the path will still be confined in longitude. One sees

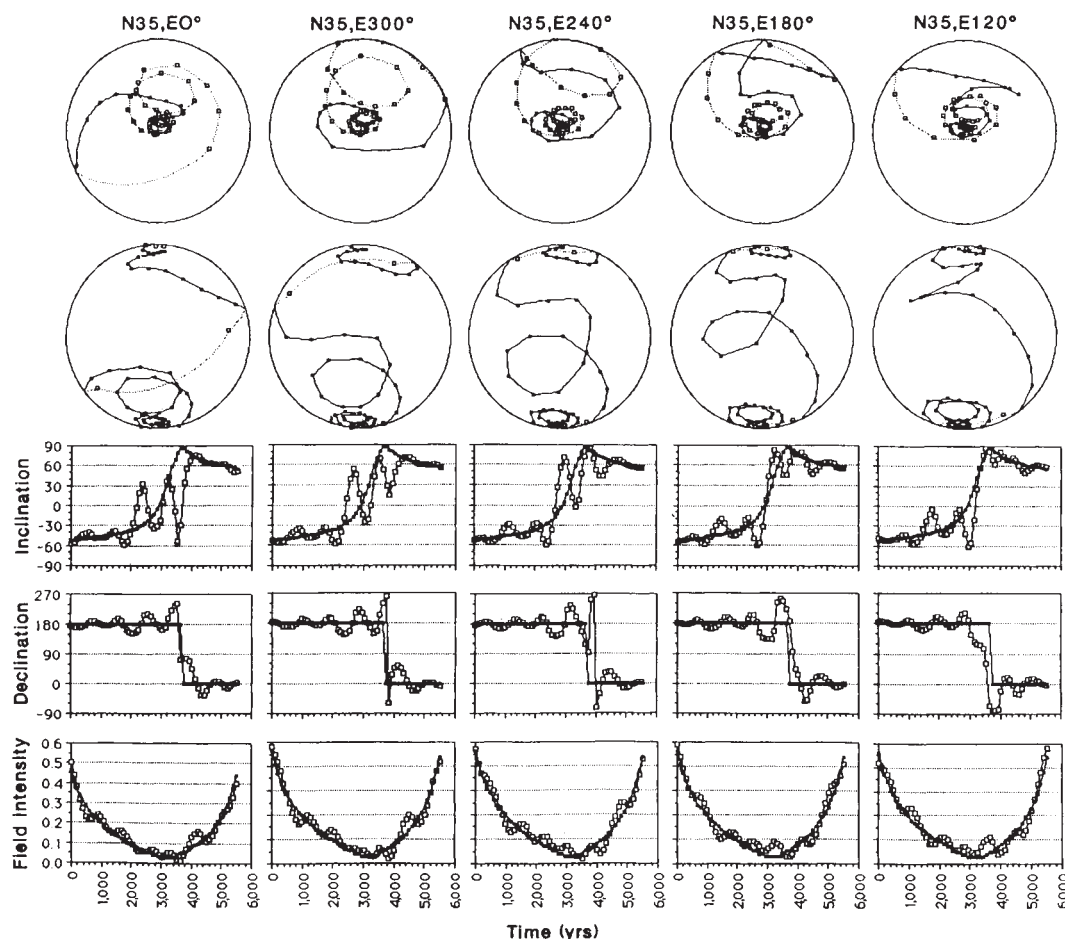


Fig. 2 Synthetic palaeomagnetic records of a reverse to normal transition of the geomagnetic field according to the zonal plus drifting non-dipole field model. Observation sites are at 35° N at longitudes 0°, 300°, 240°, 180° and 120°. The uppermost row of plots are polar plots of VGP paths. The second row gives the same VGP paths viewed from above the intersection of the equator and the site meridian. The lower profiles give inclination, declination and intensity as a function of time throughout the reversal.

the evolution of loops and hang-ups in the path as the longitude of the observation site varies. These are consequences of the phase relationships at the various sites between the growth and decay of the zonals and the passages of features of the drifting non-dipole field. The causes of individual features can be seen by comparing the intensity plots with the VGP paths for the individual sites.

The addition of a drifting non-dipole field to the earlier zonal model greatly improves simulation of the field reversal records. However, we have increased the complexity of the model. To justify this increase in complexity, we note that recent studies have indicated that secular variation may persist through reversals⁵. Indeed a much earlier attempt to model this scenario was made by maintaining an equatorial dipole throughout the reversal⁷. The present model achieves the same effect, but does not involve the use of the equatorial dipole which is precluded by the observations.

The particular model developed requires that the dipole and quadrupole families of fields interact strongly during the reversal, as energy is fed from the dipole to the quadrupole, octupole and hexadecapole. This is in conflict with the idea that these families behave independently, but appears consistent with palaeomagnetic data⁸. Such behaviour may indeed be a contrast between the reversing and non-reversing state of the field⁸. The

model is also consistent with the suggestion that the drifting non-dipole field persists throughout the reversal, and hence may be generated somewhat independently of the dipole.

We conclude that a model combining zonal harmonics with a drifting non-dipole field demonstrates that reversal records will vary with both longitude and latitude of the observation site. The form of the record is then determined by the phase relations between the zonals and the drifting features of the non-dipole field. The model appears to be able to generate many of the phenomena seen in palaeomagnetic records of reversals. Of course, such a model is not unique. Nevertheless, we suggest that during the reversal of the geomagnetic field, energy in the dipole field may be partitioned among the low-order zonals and that the drifting non-dipole field remains an important aspect of the field.

Received 13 October 1987; accepted 4 February 1988.

1. Hillhouse, J. W. & Cox, A. *Earth planet. Sci. Lett.* **29**, 51-64 (1976).
2. Hoffman, K. A. *Science* **196**, 1329-1333 (1977).
3. Hoffman, K. A. & Fuller, M. *Nature* **273**, 715-718 (1978).
4. Williams, I. & Fuller, M., *J. geophys. Res.* **86**, 11657-11665 (1981).
5. Valet, J.-P., Laj, C. & Tucholka, P. *Nature* **322**, 27-32 (1986).
6. Niitsuma, N. *Sci. Rep. Tohoku Univ. Ser. 2*, **43**, 1-39 (1971).
7. Dodson, R. et al. *Geophys. J. R. astr. Soc.* **53**, 373-412 (1978).
8. Merrill, R. T. & McElhinny, M. W. *The Earth's Magnetic Field* (Academic, New York, 1983).

Specificity of the spatial power-law exponent in ecology and agriculture

L. R. Taylor*, J. N. Perry†, I. P. Woivod‡
& R. A. J. Taylor‡

* 12 Carisbrooke Road, Harpenden, AL5 5QT, UK

† Rothamsted Experimental Station, Harpenden, AL5 2JQ, UK

‡ Ohio Agricultural Research and Development Centre, Wooster, Ohio 44691, USA

Accounting for the spatial distribution of individuals is an important issue in ecology and applied biology. The exponent, b , in the established empirical relationship¹, $s^2 = am^b$, between s^2 , the spatial sample variance and m , the mean population density, conveys information about the spatial distribution of individuals of a particular species. When randomness occurs, variance and mean are equal. Usually $b > 1$ and, for given a , individuals randomly dispersed at some low density become more aggregated as density increases at a rate associated positively with the species' b value. The frequent recurrence of the same b value in field samples of a particular species has facilitated efficient sampling programmes² for pests and wildlife. But Downing³ claimed that the index b does not differ between species, is inconsistent within species, and is statistically biased, depending substantially and systematically on sample attributes. We report here technical deficiencies in his analyses and examine more substantial data that support none of these claims.

Our original collection⁴ of 156 data sets from the literature, later supplemented to 221 sets by Downing³, was not ideal⁵ for inter-specific comparison of b values, comprising samples of selected species collected by diverse methods at varying scales. Such data lack precision when units per sample $N < 15$ and sample pairs $M < 5$. With few high-density counts, they rarely yield the high variances that occur in systematic large-scale sampling studies. R , the range of values of $\log m$ in the regression form of the power law, $\log s^2 = \log a + b \log m$, should be as large as practicable. We therefore also analysed^{5,6} three data collections comprising all species adequately represented in large, standardized samples collected simultaneously, with tens of thousands of individuals from three major taxa over many years: 111 bird, 97 aphid and 263 moth species (Fig. 1).

Although values of b for a species may change with varying conditions⁷⁻¹¹, or vary with one factor such as scale or life-stage (for instance, ref. 53 in ref. 2) while remaining constant for others¹², there is mounting evidence for consistency of b when conditions vary little. Thus, analysis of parallelism in seven independent studies of *Aphis fabae*¹³⁻¹⁷ demonstrated² a common value of b ; and b did not differ between woodland and farmland environments in 103 of 111 bird species⁵. Values of b were consistent in independent studies in refs 20, 32, 41, 43, 44, 49, 58, 60, 83, 86, 93, 96, 97, 100, 110, 129, 147, 160 and 164, listed in ref. 2. Reports continually confirm negligible differences in b for pests in agricultural systems where environmental change is minor¹⁸⁻²⁰.

Although b incorporates a species-specific component, it is not a unique taxonomic character for separating species. Overall differences between b values within collections (Fig. 1) do not imply that every b value should differ significantly from every other, as Downing presumes³. Downing's analysis (his Fig. 1) is based on the statistically doubtful²¹ procedure of multiple pairwise tests and is questionable because the standard errors (s.e.s) of $\hat{b}$ values are heterogeneous. Contrary to his conclusion that b values rarely differ, he found over 8,000 pairs of b values which differed significantly ($P < 0.1$), although less than 2,500 would be expected by chance if all species had the same b value.

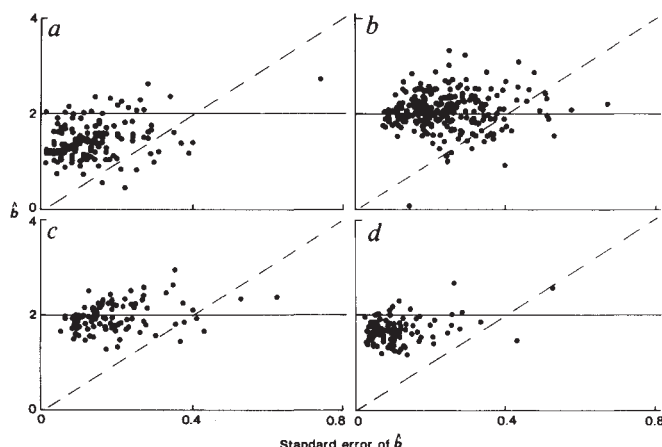


Fig. 1 Regression coefficients, $\hat{b}$, plotted against their standard errors for: a, 156 sets of data from the literature⁴; b, 263 moth species⁵; c, 97 aphid species⁵; d, 103 bird species⁵. Downing's criterion for data quality, s.e. $\hat{b}/\hat{b} < 0.2$ (dashed line), met by 78% of the regressions in his collection³, was achieved by 92%, 98%, 94% and 90% of regressions in a-d, respectively. Differences between b values are indicated when their variance is larger than that with which each is estimated; the ratio of between-to-within variance is, on average, 9.4, 4.1, 1.9 and 2.1 in a-d, respectively. If $\hat{b}$ declined substantially³ with each of N , M and R , all of which are inversely related to s.e. $\hat{b}$, trends in plots of $\hat{b}$ versus s.e. $\hat{b}$ should be readily apparent, but are absent in a, b, c and d. Claims³ that values of $\hat{b} > 2$ (solid line) are rare for regressions with $N > 60$ or $M > 20$ or $R > 1$ are unsupported by data in a-d; by the above criterion of good data, hundreds of species have $\hat{b} > 2$.

Evidence³ that intra-specific b values differ is valid only when data are comparable and adequate. Of the 19 sets in Downing's Table 1, $N < 15$ in 11 sets and $M < 5$ in six; *Bosmina coregoni* compares horizontal variability in a eutrophic lake with vertical variability in a reservoir; *Leptinotarsa decemlineata* compares adults with eggs ($\hat{b}$ is misquoted: 1.52 not 0.76); *Monodiamesa bathyphila* and *Limnodrilus hoffmeisteri* are extremely sparse data and one species determination is improbable²²; the *Popillia japonicum* samples differ in scale; *Pieris rapae* compares caged with free-flying individuals, between continents where behaviour differs²³; *Pyrausta nubilalis* compares field plots with individual plants. Also, Downing's statement³ that $\hat{b}$ is always lower where more samples are analysed is incorrect for four out of seven species. Also, the valid procedure²¹ is analysis of parallelism, not multiple t -tests performed on estimates of b with differing variances.

Downing's evidence³ of bias in $\hat{b}$ (his Fig. 2) depends on his fitting regressions to sparse data sets ($N \leq 7$, $M \leq 5$ and $R < 0.5$) that ignore the heterogeneity of s.e.s of $\hat{b}$. No such clear trends are evident in our data nor is $\hat{b} > 2$ rare as Downing states³ (Fig. 1). Also, for each of four moth species⁵ with $\hat{b} > 2$: *Agrotis segetum*, *A. exclamationis*, *Lycophotia varia* and *Ellopiia fasciaria*, the M ranked yearly means were split into high and low sets, halving M and R ; analysis of parallelism shows b and $\log a$ are not affected ($F_{2,9} = 0.44$; $F_{2,9} = 0.08$; $F_{2,12} = 2.00$; $F_{2,12} = 0.77$, respectively, $P > 0.05$), demonstrating that bias is absent.

Barbour²⁴ reiterated^{1,2} difficulties with the estimator $\hat{b}$; we had earlier discussed alternatives^{6,25,26}. Recent work¹¹ confirms that a variety of mechanisms may contribute to power-law data for a wide range of values of b , so Downing's arguments concerning models are no longer relevant.

This evidence confirms the efficiency of long-term programmes of repeated sampling based on the power law, now widely used to monitor crop pests^{2,27} and in wildlife conservation²⁸.

We thank R. M. Cormack, G. H. Freeman, J. C. Gower and M. Williamson for helpful comments. L.R.T. thanks the Leverhulme Trust for the Fellowship during the tenure of which this paper was written.

Received 7 September 1987; accepted 10 March 1988.

1. Taylor, L. R. *Nature* **189**, 732-735 (1961).
2. Taylor, L. R. *A. Rev. Entomol.* **29**, 321-357 (1984).
3. Downing, J. A. *Nature* **323**, 255-257 (1986).
4. Taylor, L. R., Woivod, I. P. & Perry, J. N. *J. anim. Ecol.* **47**, 383-406 (1978).
5. Taylor, L. R., Woivod, I. P. & Perry, J. N. *J. anim. Ecol.* **49**, 831-854 (1980).
6. Taylor, L. R. & Woivod, I. P. *J. anim. Ecol.* **51**, 879-906 (1982).
7. Wratten, S. D. *J. anim. Ecol.* **43**, 191-198 (1974).
8. Hodgson, C. J. *Ecol. Ent.* **3**, 289-298 (1978).
9. Taylor, L. R. *Can. J. Fish. aquat. Sci.* **37**, 1330-1332 (1980).
10. Taylor, L. R., Taylor, R. A. J., Woivod, I. P. & Perry, J. N. *Nature* **303**, 801-804 (1983).
11. Perry, J. N. *Oikos* **51**, 124-130 (1988).
12. Wilson, L. T. & Room, P. M. *Envir. Ent.* **12**, 50-54 (1983).
13. Johnson, C. G. *Rothamsted Exp. Stn. Report for 1949*, 77-78 (1950).
14. Banks, C. J. *Bull. ent. Res.* **45**, 751-756 (1954).
15. Taylor, L. R. *Ann. appl. Biol.* **65**, 181-189 (1970).
16. Behrendt, Von K. *Zool. Jb. Syst.* **98**, 418-454 (1971).
17. Wilding, N. & Perry, J. N. *Ann. appl. Biol.* **94**, 367-378 (1980).
18. Jones, V. P. & Parrella, M. P. *J. econ. Entomol.* **77**, 75-79 (1984).
19. Zehnder, G. W. & Trumble, J. T. *J. econ. Entomol.* **78**, 138-142 (1985).
20. Nowierski, R. M. & Gutierrez, A. P. *J. econ. Entomol.* **79**, 868-872 (1986).
21. Perry, J. N. *J. econ. Entomol.* **79**, 1149-1155 (1986).
22. Saether, O. E. *J. Fish. Res. Bd Can.* **30**, 665-679 (1973).
23. Jones, R. E. *Oecologia* **72**, 69-76 (1987).
24. Barbour, A. D. *Nature* **327**, 664 (1987).
25. Perry, J. N. *Appl. Statist.* **30**, 254-263 (1981).
26. Perry, J. N. *Appl. Statist.* **36**, 15-21 (1987).
27. Taylor, R. A. J. *Environ. Entomol.* **16**, 1-8 (1987).
28. Taylor, L. R. *J. anim. Ecol.* **55**, 1-38 (1986).

Estimation of genetic trend in racing performance of thoroughbred horses

B. Gaffney* & E. P. Cunningham†

* Genetics Department, Trinity College, Dublin 2, Ireland
† Agricultural Institute, Dublin 4, Ireland

Thoroughbred horses have been bred exclusively for racing in England since Tudor times and thoroughbred horse racing is now practised in over 40 countries and involves more than half-a-million horses worldwide. The genetic origins of the thoroughbred go back largely to horses imported from the Middle East and North Africa to England in the late seventeenth and early eighteenth centuries¹. Since the establishment of the Stud Book in 1791, the population has been effectively closed to outside sources, and over 80% of the thoroughbred population's gene pool derives from 31 known ancestors from this early period². Despite intense directional selection, especially on the male side, and the generally high heritabilities of various measures of racing performance³⁻⁵, winning times of classic races have not improved in recent decades⁶. One possible explanation for this is that additive genetic variance in performance may have been exhausted in the face of strong selection⁷. To test this, we have estimated the genetic trend in performance over the period 1952-77 using TIMEFORM handicap ratings which are based entirely on the horse's own performance, and express its racing merit as a weight in pounds which the compilers believe the horse should carry in an average free-handicap race. These ratings take into account such factors as the firmness of the ground, the distance and the level of the competition. Our results indicate that the failure of winning times to improve is not due to insufficient genetic variance in the thoroughbred population as a whole.

Our data consist of end-of-year ratings of three-year-olds, representing the best professional assessment of the merit of the horses based on their performances in that season. We have limited the study to three-year-old performances because many horses do not have a two-year-old performance rating, or retire from the track without receiving a four-year-old rating. We used a total of 31,263 three-year-old ratings representing the progeny of 2,087 stallions, for most of which a TIMEFORM rating was available. TIMEFORM ratings were also available for 59% of the dams. The means and standard deviations in TIMEFORM ratings for the sires, dams and progeny are given in Table 1.

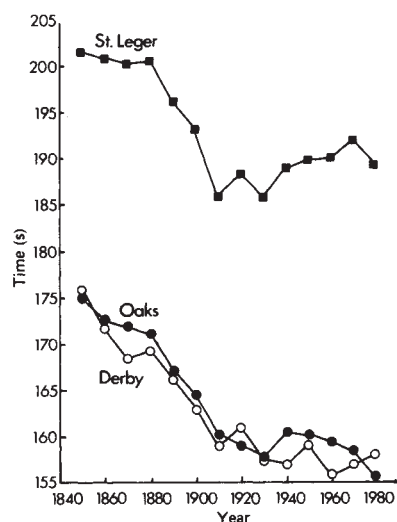


Fig. 1 Trends in winning times of the three principal classic races. Ten-year average winning times are plotted. The Oaks and Derby are 1.5-mile races for three-year-old fillies and colts respectively. The St Leger is a 1.75-mile race open to both sexes. Times were omitted for a few years when the races were run away from their usual venues. For years in which races were run over lengths different from the standard, winning times were adjusted for length of race and for change of pace associated with length.

Because there are consistent differences between the sexes, the progeny ratings are also given separately for colts, fillies and geldings. Because sires and dams are selected groups, their mean ratings are higher, and the standard deviations lower, than those of the progeny. The mean TIMEFORM ratings tend to decline over the period (regression of -0.32 ± 0.16 per year), reflecting the fact that more lowly rated horses have been included in the TIMEFORM ratings in recent years.

The distribution of progeny ratings is approximately normal (slight but significant skewness; non-significant kurtosis). If we assume normality, and also that the sires and dams are selected as an upper percentile from the progeny distribution, then it is possible to infer the selection intensity. Let the upper or selected percentile (p) begin at a value x on the abscissa, at which value the ordinate is z . The mean and standard deviation of the selected group are then $i = z/p$ and $\sigma_s = (1 - i(i - x))\sigma$. The means and standard deviations of the sire and dam ratings correspond to what would be expected if the sires represent the upper 6% of the distribution of colt ratings, and the dams the upper 52% of the filly distribution. These figures are also reasonably consistent with recruitment levels to the population as percentages of males and females born.

The average generation interval in thoroughbreds (defined as the age of the parents at the birth of their average offspring) has been estimated² as 10.1 years in the period since 1900. The average ages of sires and dams of a sample of 100 horses from the present study, born in 1981, were 11.2 ± 4.5 and 9.7 ± 3.8

Table 1 Means and standard deviations of three-year-old TIMEFORM ratings for 31,263 racehorses, for their 2,087 sires and for a proportion (59%) of their dams

	Mean	s.d.
Sires	126.63	7.70
Dams	91.34	15.78
Progeny—All	81.12	19.36
—Colts	88.89	18.98
—Fillies	78.17	18.21
—Geldings	72.53	13.73

respectively. We have used these latter values in calculating genetic change.

We estimated the heritability of TIMEFORM rating by a number of different methods: regression offspring on sire, on dam and on mid-parent, and by half-sib analysis of variance. For the regression methods, we first adjusted the data for the systematic sex differences, and expressed each rating as a deviation from the year mean. We obtained the estimate based on the half-sib analysis of variance by analysing the full data set using a linear model which included sex, year and sire.

Two potential biases in these estimates can be allowed for. First, the parents are themselves a selected group. Assuming that selection is based on TIMEFORM ratings, or on a closely correlated criterion this should not affect the regression estimates, but would affect the half-sib estimates. The appropriate adjustment⁸ has been applied. Second, bias could arise from positive assortative mating. We calculated the annual correlations between ratings of sire and dam (mean = 0.301; range = 0.250–0.358). We have applied the appropriate adjustment⁹ to the regression on sire, regression on dam and half-sib estimates. The resulting heritability estimates are presented in Table 2.

We have assumed in our analysis that covariance (offspring-parent, half-sib) is due entirely to additive genetic causes. Any additional source of positive covariance will tend to bias the estimates upwards. In the case of regression of offspring on sire, such a bias is likely: highly rated stallions have very high stud fees, and their offspring receive above-average treatment, for instance, in the quality of the trainer. Therefore, the very high

Table 2 Heritability estimates for racing performance as measured by TIMEFORM rating

Method	Heritability
Regression on sire	0.76 ± 0.023
Regression on dam	0.39 ± 0.013
Regression on mid-parent	0.47 ± 0.014
Half-sib anova	0.50 ± 0.036

estimate given by the regression of offspring on sire can be discounted. A similar bias (environmental sources of covariance between half-sibs) could also affect the half-sib estimate. The estimate based on the regression of offspring on dam is the lowest, probably reflecting the lower likelihood of environmental biases in this case. Because any bias is likely to be positive, we have assumed a slightly lower heritability value of 0.36 when predicting and estimating genetic gain.

The predicted annual rate of genetic gain is

$$h^2 \sigma_p (i_m + i_f) / (L_m + L_f) \quad (1)$$

where h^2 is the estimated heritability, σ_p the phenotypic standard deviation, i_m and i_f are the selection differentials on a standard normal distribution corresponding to the selection intensities in males (6%) and females (52%), and L_m and L_f are the generation intervals in males and females, respectively. This formula takes no account of any reduction in genetic variance due to selection in the previous generation. But the annual estimates of heritability (by all methods of estimation) were very consistent over time, suggesting that no significant reduction in genetic variance has occurred. The predicted rate of genetic change is 0.92 TIMEFORM units per annum.

The direct measurement of actual genetic gain achieved is best done by using control strains, or comparisons of divergent selection lines¹⁰. Where no genetic controls are available, genetic gain can be measured by retrospective calculation of the mean genotype of annual cohorts of the population. We obtained such an estimate by carrying out a large-scale BLUP (Best Linear Unbiased Prediction) analysis¹¹ to estimate the breeding values of the sires as expressed in the performance of their progeny, assuming a heritability value of 0.36. We included a total of 11,328 progeny of 516 sires (born 1952–77), each with three-

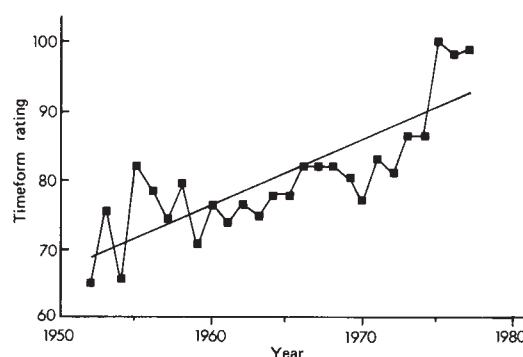


Fig. 2 Trend in estimated mean genotypes of stallions born in the years 1952–77. The regression of mean genotype on time is 0.94 ± 0.13 TIMEFORM units per year. Individual genotypes were estimated by BLUP procedures applied to end-of-year TIMEFORM ratings of three-year-old progeny. A total of 516 stallions and 11,328 progeny were involved.

year-old progeny performing in at least two of the years 1961–85. We classified sires by year of birth, and estimated the genetic change by using the mean breeding values of sires born in the different years (Fig. 2; estimated genetic trend = 0.94 ± 0.13 TIMEFORM units per year). Assuming a higher heritability value of 0.48 increased this estimated rate of genetic change by ~5%, indicating that it is not very sensitive to the heritability value used.

We calculated a second estimate of genetic change as minus twice the intra-sire regression of mean offspring performance, as a deviation from their contemporaries, on time¹². We first adjusted performances for sex differences and then expressed them as deviations from year means. This method assumes that, as a sire's genetic contribution to his offspring is constant over time, any trend in the relative performance of successive cohorts of his offspring is a reflection of genetic change in the average contribution of other sires to his offsprings' contemporaries. Bias could arise from preferential selection of mates in later stages of a stallion's stud career, based on the performance of his early crops of progeny. To avoid this effect, we used only the first four annual cohorts of progeny. The resulting estimate of genetic change (obtained as a weighted average of 585 separate intra-sire regressions of progeny performance on year in the years 1961–85) is 1.62 ± 0.26 TIMEFORM units per annum.

Of the two methods for estimating genetic change, BLUP is preferred, as it is adjusted for more sources of possible bias. This estimate of genetic trend is remarkably close to that predicted from the heritabilities, selection intensities and generation lengths. Both figures confirm that the population as a whole is improving in TIMEFORM rating at ~1% per annum. This is not inconsistent with the lack of progress in winning times of classic races. First, although both are measures of racing performance, they are not exactly the same measure, one being based on time, the other on handicap weight. Second, our estimates refer to the population as a whole, whereas the trend in winning times relates only to the best horses in the best races. We conclude that the explanation for the lack of progress in winning times is not due to a lack of genetic gain in the thoroughbred population as a whole.

We thank Mr Phil Bull for permission to use TIMEFORM data as the basis of this study.

Received 10 November 1987; accepted 2 March 1988.

1. Willett, P. *An Introduction to the Thoroughbred* (Paul, London, 1975).
2. Mahon, G. A. T. *Inbreeding and Infertility in Thoroughbred Horses* (thesis, Dublin Univ. 1980).
3. Hintz, R. L. *J. anim. Sci.* **53**, 582–594 (1980).
4. Langlois, B. *Livest. prod. Sci.* **7**, 591–605 (1980).

5. Tolley, E. A., Notter, D. R. & Marlow, T. J. *Anim. Breed. Abstr.* **53**, 163-185 (1985).
6. Cunningham, E. P. in *Proc. Int. Symp. on Genetics and Horse Breeding*, 2-8 (Royal Dublin Society, Dublin, 1975).
7. Marks, H. L. *Proc. 2nd World Congress on Genetics Applied to Livestock Production*, Madrid **5**, 84-92 (1982).
8. Ponzoni, R. W. & James, J. W. *Theor. appl. Genet.* **53**, 25-27 (1978).
9. Falconer, D. S. *Introduction to Quantitative Genetics* (Layner, London, 1981).
10. Hill, W. G. *Animal Breeding Abstracts* **40**, 1-15 (1972).
11. Henderson, C. R. *Biometrics* **31**, 423-447 (1975).
12. Smith, C. *Anim. Prod.* **4**, 239-251 (1962).

Direct observation of microtubule dynamics in living cells

Paul J. Sammak & Gary G. Borisy

Laboratory of Molecular Biology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

The study of cell locomotion is fundamental to such diverse processes as embryonic development, wound healing and metastasis. Since microtubules play a role in establishing the leading lamellum and maintaining cell polarity¹⁻⁵, it is important to understand their dynamic behaviour. *In vitro*, subunits exchange with polymer by treadmilling^{6,7} and by dynamic instability^{8,9}. Disassembly events can be complete¹⁰ (catastrophic) or incomplete¹¹ (tempered). *In vivo*, microtubules are in dynamic equilibrium with subunits¹² with a half-time for turnover of 4-20 min¹³⁻¹⁵. Microtubules grow by elongation of their ends^{14,16} and are replaced one by one with turnover being most rapid at the periphery¹⁵. Although previous results are consistent with dynamic instability, we sought to directly test the mechanism of turnover. Direct observations of fluorescent microtubules in the fibroblast lamellum show that individual microtubules undergo rounds of assembly and disassembly from the same end. Reorganization of the microtubule network occurs by a tempered mode of dynamic instability.

To observe microtubule dynamics, single microtubules must be detected with adequate clarity. To confirm that we could detect single fluorescent microtubules, *in vitro* preparations were observed by both light and electron microscopy. Figure 1a shows microtubules assembled from a mixture of 5- and 6-carboxy-Xrhodamine-N-hydroxysuccinimide-labelled tubulin (Xrh-tubulin)¹⁷ and unlabelled tubulin. The figure clearly shows single, unbranched fibres, and electron microscopy confirmed that the preparations consisted predominantly of single microtubules¹⁷.

Single microtubules were also detected in the lamellum of living cells (Fig. 1b). The network of fibres overlapped, but individually the fibres were unbranched, of constant width, and were comparable in brightness with microtubules *in vitro*. Electron microscopy has shown that most fibres detected by anti-tubulin immunofluorescence in fibroblast lamella are single microtubules¹⁸, and we confirmed in our preparations that direct Xrh-microtubule fluorescence corresponded to anti-tubulin immunofluorescence¹⁷. The diameter of microtubules was 0.3 μm (near the diffraction limit), and when two microtubules overlapped, the zone of overlap was clearly brighter. Thus we could distinguish microtubules from bundles by criteria of width, brightness and morphology.

Living cells were observed over time to measure changes in position of the microtubules. Figure 2 shows three sequential images of the lamellum shown in Fig. 1b. Comparison of these images demonstrates directly that individual microtubules lengthen, shorten and move laterally in the lamellum of living fibroblasts.

Although individual microtubules appear to shorten, they might have disassembled at the plus end or have been pulled

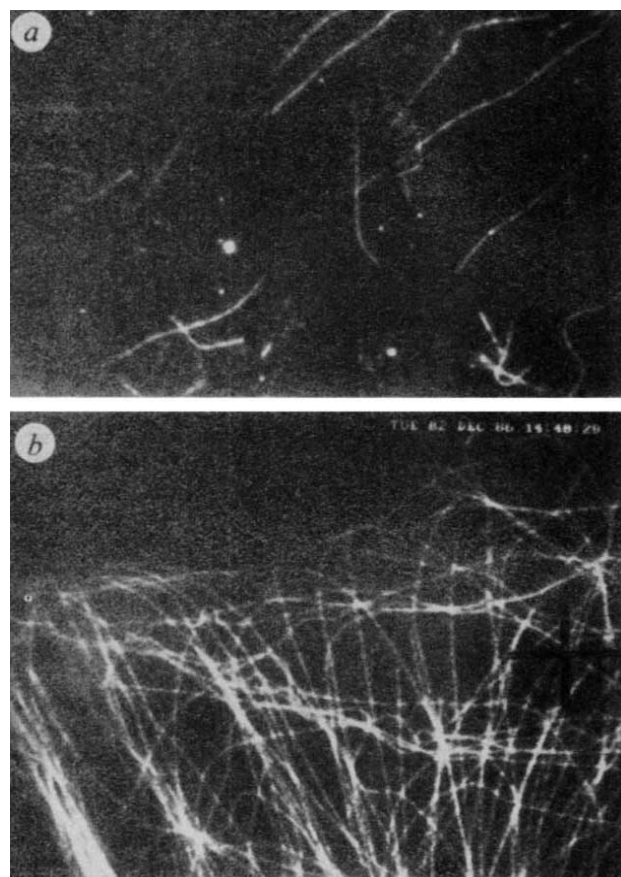


Fig. 1 a, Single microtubules *in vitro* detected by digital video microscopy. b, Individual microtubules in the lamellum of a living fibroblast. Fluorescence observations were made with a $\times 100$ planapo objective on a Zeiss IM-35 (Thornwood, New York)¹⁷. Video frames (30 s^{-1}) from a SIT camera (Dage-MTI Model 66) were digitized and summed with an image processor (Quantex Corp., Model 9200) to reduce noise²². Images were background subtracted to correct for camera shading and contrast was enhanced with look-up tables. Image in (a), 18 frame sum. Image in (b), 28 frame sum. Single microtubules were identified as uniformly fluorescent, unbranched fibres. Areas of microtubules overlap appeared brighter. Cross hairs, 10 μm . Magnification and brightness of microtubules are similar in (a) and (b).

Methods. a, Tubulin was labelled with Xrhodamine¹⁷ to produce a dye to protein ratio of 0.43 (extinction coefficient, 63,000 in PIPES at pH 6.9). Protein yields after derivatization and two cycles were typically 40-50%. The effects of Xrh-tubulin on the assembly of tubulin were minimal as determined by an *in vitro* seeded assembly assay; for a mixture with the same dye-to-protein ratio as injected cells (0.09), the initial rate and the final extent of polymerization were 96% and 87% of control values, within the variation of repeated controls¹⁷. b, Human neonate foreskin fibroblasts, line 356, were cultured, microinjected, and temperature-cycled as previously reported¹⁵ to produce uniformly-labelled microtubules. To evaluate whether labelled microtubules behaved differently from unlabelled ones *in vivo*, different preparations of Xrh-tubulin were microinjected to produce final cellular dye to protein ratios estimated to range between 0.02 and 0.2. The frequency and length of microtubule excursions per cell were categorized as either more active or less active (see activity threshold in text). Over the range examined, no effect of labelling stoichiometry was detected.

in towards the cell centre. Any mechanism of pulling, including treadmilling^{6,7} or sliding¹⁹, could be tested by placing a reference mark on each microtubule. In Fig. 3, fluorescent microtubules were marked by photobleaching with laser light. A pulling model would predict that the distance between the mark and the distal end should always remain constant. In each case, however, the

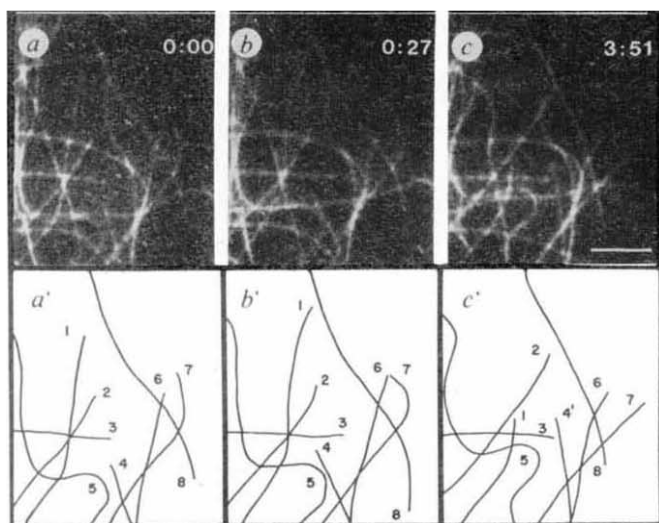


Fig. 2 Individual microtubules undergo length excursions. Cells were injected with Xrh-tubulin and imaged as described for Fig. 1 except that no background subtraction was performed. Length changes were determined with respect to the bulk of cytoplasmic microtubules which defined the reference frame and were measured by superimposing aligned images on the video monitor and toggling between them. Panels *a'*, *b'* and *c'* are tracings of images *a*, *b* and *c*, showing selected microtubules. Microtubule 2 grew, microtubules 1, 6, 7 and 8 grew and then shrank, and microtubule 3 remained approximately the same length. Microtubule 4 appears to grow and move laterally, but an intermediate image (not shown) revealed that it shrank to the edge of the panel and subsequently regrew. Some microtubules (5, 7) underwent substantial lateral movements. Ambiguous microtubules were frequently resolved by detailed comparisons among a series of images. For example, a horizontal domain of microtubule 5 (*b'*) is distinct from its neighbours in panel *a* and *c* but overlaps them in *b* where the intensity is higher. Microtubules that could not be unambiguously determined were omitted from subsequent analysis. Bar, 3 μm . Each image, 15 frame sum. Elapsed time (min : s) is shown in this and successive figures in the upper right hand corner.

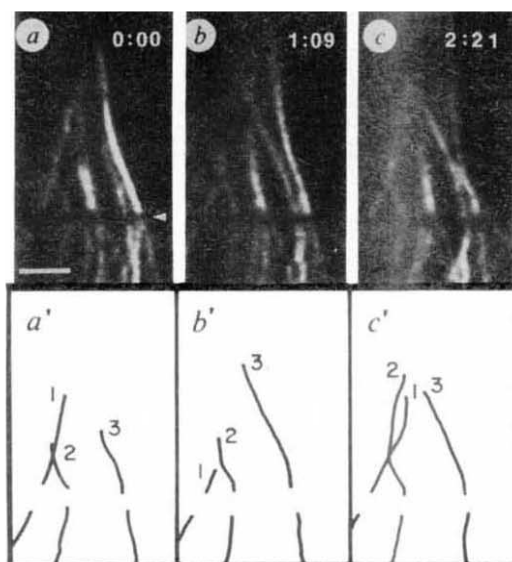


Fig. 3 Growing and shortening *in vivo* are due to subunit addition and loss at the plus end. Cells were injected with Xrh-tubulin and imaged as in Fig. 2. A reference bar (arrowhead in *a*) was produced across the microtubules by photobleaching the dye with the 514 nm line of an argon ion laser. Anti-tubulin immunofluorescence of cells bleached with the same energy showed microtubules were not broken by the bleaching^{15,17}. Panels *a-c* were aligned so that the bleach zone (and hence the microtubules) were in the reference frame. Panels *a'-c'* were made from tracings of panels *a-c*, and only microtubules with unambiguous ends are illustrated. The end of microtubule 1 shrank and then grew with respect to its bleached domain, microtubule 2 grew, and microtubule 3 grew and then shrank. Length changes were due to gain and loss of subunits distal to the bleached domain on each microtubule. Bar, 3 μm . Each image, 15 to 18 frame sums.

distance between the microtubule end and its mark decreased or increased. In addition, since the bleached domains of these microtubules remained non-fluorescent while competent labelled subunits added to the ends, mechanisms of label incorporation such as intercalation into the microtubule wall or anomalous wall binding can be excluded. We conclude that the length excursions observed were due to rounds of gain and loss of subunits at the plus end.

Although length excursions clearly occur *in vivo*, the question remained whether their magnitude and frequency are sufficient to produce turnover in the lamellum. Figure 4 shows the excursions of eight microtubules in one lamellum over a 15 min interval. In this lamellum nearly 90% of the microtubule domains turned over within 7 min which corresponds to a half-time of 2 to 3 min, in agreement with half-times measured by fluorescence replacement after photobleaching^{13,15}.

Microtubules behave independently over the short term as neighbouring microtubules did not shrink and grow in unison. Microtubules behaved diversely over the long run as some underwent long, uninterrupted excursions, some underwent repeated rounds of growing and shrinking, and others showed little net change in length. Most shortening excursions were followed by growth and did not disassemble the entire microtubule. This tempered instability¹⁵ produced a variety of behaviour in terms of the magnitude and frequency of excursions and was sufficient to account for the turnover of microtubules in the lamellum.

Most lamella exhibited these microtubule excursions, although some cells were more active than others. An arbitrary

threshold of activity was defined as three 5 μm excursions per min per 1,000 μm^2 of lamellum. By this definition, 67 of 88 cells observed were active and 21 were either too congested to score or exhibited very short or very few excursions.

Most microtubules showed tempered instability. Of 96 shortening excursions analysed, 51% were followed by conversions to the growing phase. This number is an underestimate as it includes some microtubules that were only observed for one time interval. Fifty-three shrinking microtubules were observed for at least two time intervals and of these, 81% converted to the growing phase. Thus, we estimate that between 50 and 80% of the shortening excursions were tempered, and conversely that 20 to 50% might have been catastrophic.

Dynamic instability *in vitro* is characterized by slow frequent growth and fast infrequent shortening and our observations show that this is also the case *in vivo*. Among 390 excursions of 98 microtubules in 4 cells, the mean growth rate, $3.5 \pm 3.2 \mu\text{m min}^{-1}$, was slower than the mean shrinking rate, $4.3 \pm 5.7 \mu\text{m min}^{-1}$. The maximum rates observed were +19 and -36 $\mu\text{m min}^{-1}$ respectively. The mean time interval between images, 1.4 ± 0.8 min, probably exceeded the period of many excursions. Growing and shrinking would therefore be combined in some measurements and the measured rates and excursions are probably underestimates of the mean instantaneous values (see Fig. 2). The averaged values are more appropriate for determining the mechanism of turnover since turnover is much slower than the observation intervals. Of the microtubule excursions sampled, 50% were positive, 32% were negative, and 18% were less than $\pm 0.5 \mu\text{m min}^{-1}$ and were scored as being static. This means that microtubules spend 50% of the time growing, 32% of the time shrinking, and 18% of the time pausing. We conclude that microtubules spend more time growing than shortening and that growth is slower than shortening

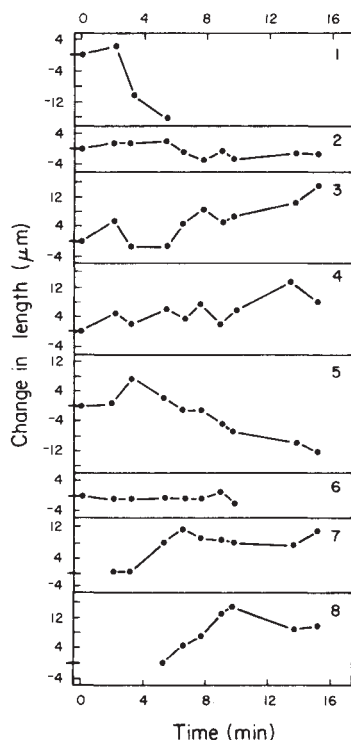


Fig. 4 Adjacent microtubules independently undergo rounds of assembly and disassembly. The change in length of 8 microtubules in one lamellum are plotted versus time of observation. Images were taken periodically over 15 min. Length changes measured as in Fig. 2. Microtubule 1 monotonically disassembled 18 μm until lost in a bundle; microtubule 8 grew 16 μm until reaching the cell edge. Microtubule 4 underwent 4 transitions between assembly and disassembly; microtubule 6 paused as if there was a cessation of length changes or a balance of very short fluctuations. All microtubules but 1 and 6 showed tempered instability during the period of observation, that is the shortening phase was followed by a growing phase and did not lead to complete disassembly of the microtubule.

as predicted by dynamic instability. Thus, dynamic instability describes the behaviour of individual microtubules *in vivo*, and is sufficient to account for turnover of the network in fibroblast lamella.

The coexistence of growing and shrinking microtubules is the basis for dynamic instability, but unlike the original formulation^{8,9}, we found that conversions between the shrinking and growing phases were not infrequent. Conversions occurred with a minimum frequency of about 1 per min per microtubule. What is similar in both the current and the original formulation, is infrequent complete disassembly. This generally tempered, as opposed to catastrophic, instability was a characteristic feature of microtubule dynamics in the lamellum and accounts for more rapid turnover in the periphery than in the cell centre¹⁵. We calculate that after shortening about 5 μm , a microtubule is equally likely to regrow as it is to continue to shorten. This value falls within the range of *in vitro* measurements of 2.4 μm ¹¹ and theoretical estimates of 20 μm ^{9,20}.

The conversion from shrinking to growing might be under cellular regulation and could be important in the control of microtubule organization. Frequent conversion would produce tempered instability, stabilizing domains in the cell centre. Infrequent conversion would produce catastrophic instability, rapidly turning over the entire microtubule array. Conversions might be different throughout the cell cycle and throughout development.

What is the biological significance of tempered instability? One possibility is that it provides a mechanism for efficiently

accelerating the pace of turnover in the lamellum without requiring rearrangement of the entire network. Greater plasticity of peripheral microtubules could explain how the network rearranges quickly as cells reorient their advancing lamellum^{5,21}. Thus tempered instability could be an important point of control of cell motility and morphogenesis.

This work was supported by the NIH.

Received 17 July 1987; accepted 29 February 1988.

1. Schliwa, M. *The Cytoskeleton* (Springer, Vienna, 1986).
2. Kirschner, M. & Mitchison, T. *Cell* **45**, 329-342 (1986).
3. Vasiliev, J. M. & Gelfand, I. M. in *Cell Motility* (eds Goldman, R., Pollard, T. D. & Rosenbaum, J.) 279-364 (Cold Spring Harbor Laboratory, New York, 1976).
4. Gottlieb, A. I., May, L. M., Subrahmanyam, L. & Kalnins, V. I. *J. cell Biol.* **91**, 539-594 (1981).
5. Kupfer, A., Louvard, D. & Singer, S. J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2603-2607 (1982).
6. Margolis, R. L. & Wilson, L. *Cell* **13**, 1-8 (1978).
7. Farrell, K. W., Jordan, M. A., Miller, H. P. & Wilson, L. *J. cell Biol.* **104**, 1035-1046 (1987).
8. Mitchison, T. & Kirschner, M. *Nature* **312**, 232-237 (1984).
9. Mitchison, T. & Kirschner, M. *Nature* **312**, 237-242 (1984).
10. Kristofferson, D., Mitchison, T. & Kirschner, M. *J. cell Biol.* **102**, 1007-1019 (1986).
11. Horio, H. & Hotani, H. *Nature* **321**, 605-607 (1986).
12. Inoue, S. & Sato, H. *J. Gen. Physiol.* **50**, 259-292 (1967).
13. Saxton, W. M. *et al. J. cell Biol.* **99**, 2175-2186 (1986).
14. Schulze, E. & Kirschner, M. *J. cell Biol.* **102**, 1020-1031 (1986).
15. Sammak, P. J., Gorbisky, G. J. & Borisy, G. G. *J. cell Biol.* **104**, 395-405 (1987).
16. Soltys, B. J. & Borisy, G. G. *J. cell Biol.* **100**, 1682-1689 (1985).
17. Sammak, P. J. & Borisy, G. G. *Cell Mot. Cytoskel.* **9** (in the press).
18. Osborn, M. & Weber, K. *Meth. Cell Biol.* **24**, 97-131 (1982).
19. Koonce, M. P., Tong, J., Euteneuer, U. & Schliwa, M. *Nature* **328**, 737-739 (1987).
20. Chen, Y. & Hill, T. L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4127-4131 (1985).
21. Aberchrombie, M. *Proc. R. Soc. B207*, 129-147 (1980).
22. Inoué, S. *Video Microscopy* (Plenum, New York, 1986).

Dispersal and the sex ratio at birth in primates

C. N. Johnson*

Department of Zoology, The Australian National University,
GPO Box 4, Canberra ACT 2601, Australia

The females of many species of primates settle for life within the home ranges of their mothers, whereas males disperse as immatures¹. According to the theory of sex allocation, the costs incurred by mothers through local competition for resources with their philopatric daughters should favour the evolution of male-biased sex ratios at birth^{2,3}. I report here two tests of this hypothesis based on data from 15 genera of primates. First, I show that the intensity of competition for resources within kin groups is strongly and positively correlated with sex ratios at birth. Second, I show that sex ratios at birth are higher in genera with female-biased philopatry than in genera in which philopatry is not female-biased. These analyses suggest that local resource competition among kin powerfully influences the evolution of sex ratios in primates.

The reproductive success of females in several species of primates is limited by the availability of food⁴, and the observation that reproductive success generally declines with group size suggests that females in groups compete with one another for access to resources^{5,6}. The regression coefficient of reproductive success on group size provides an index of the intensity of within-group competition, which can be used in intergeneric comparisons. The hypothesis of local resource competition predicts that where this index takes high values, and where groups are composed of related females, sex ratios at birth should be male-biased.

I found equations relating reproductive success of females to group size for all of the seven genera of primates in which the appropriate information on female reproductive success and sex ratios at birth is available, and in which females are known to be more philopatric than males. I made comparisons between

* Present address: Queensland National Parks and Wildlife Service, P.O. Box 1395, Rockhampton, Queensland 4700, Australia.

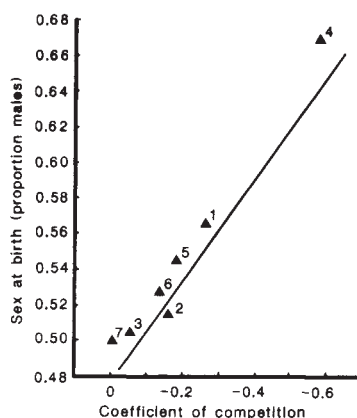


Fig. 1 Relation between the intensity of competition among females (coefficient of the regression of number of immatures per adult female on number of adult females per group) and sex ratio at birth in genera of primates with female-biased philopatry ($y = -0.29x + 0.48$, $\chi^2 = 17.63$, $P < 0.01$). I calculated statistics separately for each species, and then combined them, by averaging, into genera. The logistic regression weights each point according to sample size¹¹. The genera (with species represented, and total sample sizes used to calculate sex ratios, in parentheses) are (1) *Lemur* (*catta* 1,188), (2) *Macaca* (*mulatta*, *sylvanus*, 3,220), (3) *Papio* (*anubis*, *cynocephalus*, *ursinus*, 864), (4) *Cercopithecus* (*aethiops*, 100), (5) *presbytis* (*entellus*, 143), (6) *Colobus* (*guereza*, 317), (7) *Erythrocebus* (*patas* 479). Information collated from ref. 19 and sources cited therein.

genera because most variation in ecological and life-history traits in primates resides at that level, or at the level of the sub-family^{7,8}; sample sizes are too small to permit comparisons above genus level. I defined philopatry as residence by breeding adults within the natal home range. I measured reproductive success from census data as the numbers of immatures (infants and juveniles) per adult female. I then calculated lines of best fit to the plots of reproductive success against the numbers of adult females in groups using reduced major-axis analysis⁹.

I calculated sex ratios from zoo births recorded in the International Zoo Yearbook between 1965 and 1983. Two observations suggest that sex-ratio biases in the wild recur in zoo populations: (1) an extremely male-biased sex ratio at birth in wild *Galago*³ corresponds approximately with an excess of males born in zoos (male proportions of 0.77 ($n = 17$) and 0.61 ($n = 674$), respectively); and (2) an extremely female-biased sex ratio at birth in wild *Ateles*¹⁰ is likewise reflected in zoo births (0.27 ($n = 44$) and 0.36 ($n = 104$), respectively). In the few other cases where comparisons of sex ratios at birth in wild and captive primates can be made, they tend to be close to parity.

There is a strong relation between intensity of competition among females and the sex ratio at birth in this sample of genera (logistic regression¹¹, $\chi^2 = 17.628$, d.f. = 1, $P < 0.001$) (Fig. 1). I found no significant correlations when the measures used to calculate the index of competition within groups (group size, and the ratio of immatures to adult females) were each regressed separately against sex ratios at birth.

Two other hypotheses applicable to primates suggest that sex ratios at birth should vary between genera. The first of these predicts a positive relation between the ratio of adult male to adult female weights, and the sex ratio at birth. This is predicted because, in those mammal species in which adult males are larger than adult females, the mortality rates of juvenile males are high by comparison with those of juvenile females¹². This differential mortality of juvenile males must reduce the average cost to the parent of each male born, and should therefore select for correspondingly male-biased sex ratios at birth^{13,14}. Alternatively, if size dimorphism of adults is achieved at least partly via sex differences in early growth rates, and if members of the

faster-growing sex make heavier energetic demands on their mothers, a negative relation between sexual dimorphism of adults and the sex ratio at birth is expected.

Sexual dimorphism and sex ratios at birth are not significantly related to one another across the seven genera analysed in Fig. 1 (logistic regression, $\chi^2 = 2.56$, n.s.). This suggests either that the two possible effects of sexual dimorphism on sex ratios at birth cancel one another, or that neither of them is important in primates.

A second prediction of the hypothesis of local resource competition is that sex ratios at birth should tend to be higher in genera with female-biased philopatry than in genera in which philopatry is not female-biased. A simple enumeration and comparison of biases is not, however, an adequate way to document such an effect, because of the possibility that sex ratios at birth might vary with ecological circumstances, which could in turn differ between the two classes of genera. In particular, a negative relation between the numbers of females using overlapping home-ranges and the sex ratio at birth across genera with female-biased philopatry is expected. This is because when mothers and daughters share home ranges which do not overlap with those of less closely related females, all direct competitive interactions between females will involve close kin, and selection for male-biased sex ratios at birth will be strong. But as the number of females using overlapping home-ranges increases (as groups become larger, or as the extent of overlap between neighbouring groups increases), the effects of competition between closely related females will be outweighed by competition from females of different matriline or groups, and selection for male-biased sex ratios will weaken. The expected difference in sex ratios at birth between genera with and without female-biased philopatry can, therefore, be stated more precisely as follows: sex ratios at birth should be higher in genera with female-biased philopatry than in other genera when the effects of home-range overlap among females on sex ratios at birth are taken into account.

This prediction is confirmed in Fig. 2, which relates sex ratios at birth to home range overlap in the two classes of genera. Home-range overlap is defined as the number of females living within an area the size of the average female's home range, and is calculated as the average female home range size multiplied by the female

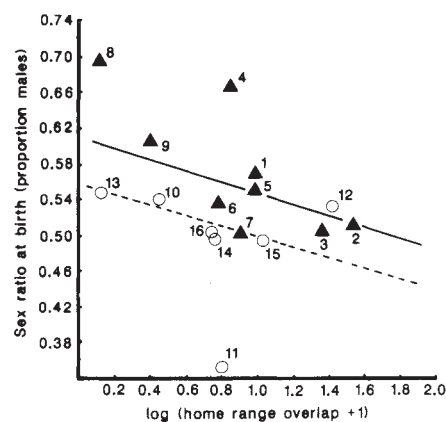


Fig. 2 Relation between log (female home-range overlap + 1) and sex ratios at birth for primate species with female-biased ($\blacktriangle$, and full line) and non-female-biased philopatry ($\circ$, and dashed line). The lines of best fit have intercepts of 0.61 and 0.55, and a common slope of -0.06 ($\chi^2 = 17.23$, $P < 0.002$). Genera are those in Fig. 1, with the addition of: (8) *Perodicticus* (*potto*, 33), (9) *Galago* (*crassicaudatus*, *demidovii*, *senegalensis* (674), (10) *Saguinus* (*oedipus*, 616), (11) *Ateles* (*paniscus*, 104), (12) *Papio* (*hamadryas*, 382), (13) *Hylobates* (*lar*, *syndactylus*, 559), (14) *Pongo* (*pygmaeus*, 575), (15) *Pan* (*trogodytes*, 891), (16) *Gorilla* (*gorilla*, 261). Information from ref. 19 and sources cited therein.

density¹⁵. Sex ratios at birth and home-range overlap are significantly and negatively related to one another across all genera (logistic regression, $\chi^2 = 10.61$, d.f. = 1, $P < 0.01$). But a model in which sex ratios at birth are assumed to be consistently higher in genera with female-biased philopatry provides an improved fit to the data ($\chi^2 = 17.23$, d.f. = 1, $P < 0.001$). The data do not justify a model in which the slopes as well as the elevations of the relations differ between the two classes of genera ($\chi^2 = 3.26$, d.f. = 1, n.s.).

This analysis provides strong evidence of the influence of local resource competition on the evolution of sex ratios in primates. By implication, it also supports less direct evidence of relations between intrasexual competition and sex ratios at birth in deer¹⁶ and marsupials^{17,18}. Also, the observation that, in primates, extreme sex ratios arise only when competition between female kin is intense (or when sex-differences in dispersal are strikingly reversed¹⁰) suggests that local resource competition overwhelms other effects which could in theory influence interspecific variation in their sex ratios.

I thank Dorothy Cheney, Tim Clutton-Brock, Andrew Cockburn, Robert Heinsohn and Peter Jarman for their comments on the manuscript. I am also grateful to Andrew Cockburn

and Ross Cunningham for statistical advice. This research was supported by an Australian Research Grants Scheme grant to Andrew Cockburn.

Received 5 October 1987; accepted 8 March 1988.

1. Pusey, A. E. & Packer, C. in *Primate Societies* (eds Smuts, B. et al.) 250-266 (University of Chicago Press, 1987).
2. Bulmer, M. G. *Heredity* **56**, 69-73 (1986).
3. Clark, A. B. *Science* **20**, 163-165 (1978).
4. Silk, J. B. in *Primate Societies* (eds Smuts, B. et al.) 318-329 (University of Chicago Press, 1987).
5. van Schaick, C. P. *Behaviour* **87**, 120-144 (1983).
6. Janson, C. *Behav. Ecol. Sociobiol.* **18**, 125-138 (1985).
7. Clutton-Brock, T. H. & Harvey, P. H. *J. Zool., Lond.* **183**, 1-39 (1977).
8. Harvey, P. H. & Clutton-Brock, T. H. *Evolution* **39**, 559-581 (1985).
9. Harvey, P. H. & Mace, G. M. in *Current Problems in Sociobiology* (ed. King's College Sociobiology Group) 343-362 (Cambridge Univ. Press, U.K., 1982).
10. MacFarland Symington, M. *Behav. Ecol. Sociobiol.* **20**, 421-426 (1987).
11. Dobson, A. J. *An Introduction to Statistical Modelling* (Chapman and Hall, London, 1983).
12. Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. *Nature* **313**, 131-133 (1985).
13. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
14. Clutton-Brock, T. H. & Albon, S. D. in *Current Problems in Sociobiology* (ed. King's College Sociobiology Group) 223-248 (Cambridge University Press, 1982).
15. Damuth, J. *Biol. J. Linn. Soc.* **15**, 185-193 (1981).
16. Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer: Behaviour and Ecology of Two Sexes* (Cambridge University Press, 1982).
17. Cockburn, A., Scott, M. P. & Dickman, C. R. *Oecologia* **66**, 427-429 (1985).
18. Johnson, C. N. *Behav. Ecol. Sociobiol.* **19**, 143-150 (1986).
19. Smuts, B. et al. *Primate Societies* (University of Chicago Press, 1987).

A group specific anamnestic immune reaction against HIV-1 induced by a candidate vaccine against AIDS

Daniel Zagury*, Jacky Bernard*, Remi Cheyner*, Isabelle Desportes*, Regine Leonard*, Michelle Fouchard*, Brigitte Reveil*, Daniele Ittele*, Zirimwabagangabo Lurhuma†, Kalumbu Mbayo†, Justin Wane†, Jean-Jacques Salaun‡, Bernard Goussard‡, Loic Dechazal‡, Arsene Burny§, Peter Nara|| & Robert C Gallo||

* Institut Jean Godinot, Université P. et M. Curie, 4 Place Jussieu, Paris, France

† Cliniques Universitaires, Kinshasa, Zaïre

‡ Institut National de Recherches Biomédicales, Kinshasa, Zaïre

§ Département de Biologie Moléculaire, Université Libre de Bruxelles, Rhode Sainte Genese, 1640 Belgique

|| National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

The first experimental immunization of humans against the AIDS retrovirus, HIV-1, was started in a series of HIV seronegative, healthy volunteers in November 1986¹. For the primary vaccination recombinant vaccinia virus (V25)² expressing the complete gp160 *env* protein³ of the HTLV-III_B strain^{4,5} of HIV-1 was introduced by scarification. This elicited a weak primary response which we subsequently attempted to enhance by additional immunizations (boosting), using four different immunization protocols. We report here that intravenous injection of paraformaldehyde-fixed autologous cells infected *in vitro* with V25 (individual D.Z.) gave the best results. This individual received second and third boosts of intramuscular gp160 derived from an HTLV-III_B clone using the hybrid vaccinia virus/bacteriophage T7 expression system⁶. An anamnestic humoral and cellular immune reaction was achieved for over one year after the original vaccination, with high levels of antibodies to the viral envelope, and neutralizing antibodies against divergent HIV-1 strains such as HTLV-III_B^{4,5,7} and HTLV-III_{RF} (also called HTLV-III HAT)^{3,5} after the first boost. In addition, group-specific cell-mediated immunity and cell-mediated cytotoxicity against infected T4 cells were obtained after the primary vaccine and enhanced by the boosts. Finally, skin tests showed both immediate and delayed hypersensitivity to gp160 *in vivo*. Although this protocol is not practical for a large scale

vaccine trial, our results show for the first time that an immune state against HIV can be obtained in man.

A vaccinia virus recombinant (V25)^{2,3} expressing gp160 *env* determinants of HTLV-III_B at the surface of infected cells was used for priming a series of HIV-seronegative healthy individuals. The rationale of our approach, based on our previous studies of HIV-1 infected cells^{8,9}, was to trigger cell-mediated immunity (CMI) against HIV through cell membrane gp160 signals. In a preliminary report we described evidence for a significant though weak immune reaction after V25 priming¹. These results prompted us to boost the primary immune response. We used four different protocols: slow drip intravenous (i.v.) infusion with paraformaldehyde fixed autologous cells infected with V25 (individual D.Z.), repeated scarification with V25 (four individuals), scarification with a fragment of gp120 *env* protein (Pgp120) which reacted with an anti-gp120 MAb (six individuals), and intramuscular injection of purified autologous cell membrane infected with V25 (individual B.G.). After boosting with either V25 vaccinia-recombinant, Pgp120 or membranes from autologous cells infected with V25, group-specific CMI persisted but was not significantly enhanced in comparison with CMI obtained after V25 priming alone¹. In addition, although most sera from individuals boosted with Pgp120 or V25 showed Western blot bands at gp160, gp120 or gp41, their neutralizing antibody titres against HTLV-III_B and HTLV-III_{RF} strains were low and variable (not exceeding 50) 1-4 months after boosting. Serum from individual B.G. showed no evidence of antibodies by Western blot but did show transient low neutralizing activity against both B and RF strains. By contrast, boosting with V25-infected fixed cells (D.Z.) provided strong humoral and cellular group specific anamnestic immune responses that persisted for at least one year. We report here the conditions and the results of this immunization.

Priming of D.Z. was done by scarification using 10⁷ plaque-forming units (PFU) of V25. After 102 days, 3.5 × 10⁷ autologous activated peripheral blood mononuclear cells infected with V25 and fixed with paraformaldehyde were administered by i.v. infusion. On days 299 and 350, 100 µg and 600 µg of HTLV-III_B gp160, purified from Vero cells co-infected with recombinant vaccinia viruses vTF7-3 and vPE-5⁵, were given intramuscularly as additional vaccine boosts. Two seronegative non-immunized volunteers received the same purified gp160 in parallel. As expected, the latter individuals did not exhibit any humoral or cellular immune response, but the vaccinated individual did, as described below.

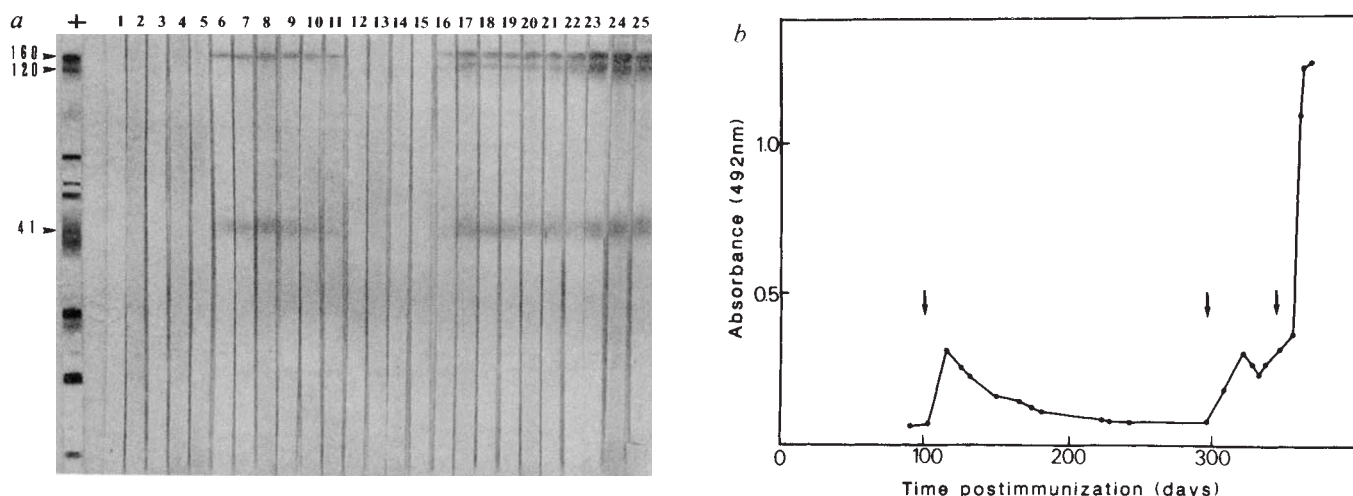


Fig. 1 Analysis of D.Z. sera collected before and during the course of immunization for antibodies to HIV. *a*, Immunoblotting with serum samples collected 0, 21, 52, 89, 102, 115, 125, 132, 150, 166, 174, 181, 223, 227, 242, 308, 321, 328, 333, 336, 347, 355, 361, 363 and 367 days after the primary vaccination (lanes 1 to 25 respectively). Lane (+), HIV positive control serum. The first boost (with fixed autologous V25 infected peripheral blood cells) was on day 102. The second and third boosts (with intramuscular gp160) were on days 299 and 350, respectively. *b*, Kinetic analysis of D.Z. sera for antibodies to HIV envelope antigens by an enzyme immunoassay. Arrows indicate the time of each of three immunization boosts performed as described in the text. The control D.Z. serum was collected prior to any immunization (5 November 1986), and it did not show any evidence of antibody (not shown).

Methods. The enzyme immunoassays were performed as described in the commercial test procedure (Abbott). Beads coated with recombinant DNA-derived HTLV-III_B core and envelope antigens were incubated 30 min. With the serum samples. After aspiration of unbound material and washing of the beads, goat antibody to human IgG conjugated with horseradish peroxidase (HRPO) was added to 30 min incubation. Unbound enzyme conjugate was aspirated, the beads washed and *O*-phenylenediamine solution containing hydrogen peroxide was added for 30 min. The reaction with HRPO yielded a yellow-orange colour; its intensity, which is in proportion to the amount of anti-HTLV-III in the sample, was measured by the Optical absorbance at 492 nm of the bead supernatants.

The humoral response to HIV was studied by Western blotting (WB), enzyme-linked immunosorbent assay (ELISA), radio-immunoprecipitation (RIPA) and virus-neutralizing activity. As shown in Fig. 1*a*, lane 6, antibody against gp160 was first observed by WB on day 115, 13 days after the first boost and lasted for about 3 months. No antibody against gp120 was found in this period, but a faint reactivity to gp41 could be seen. Nine days after the second boost, however, (lane 16, Fig. 1*a*), serum Ab reacting to both gp160 and gp120 was seen together with a faint anti-gp41 reaction and was strongly enhanced by the higher dose of gp160 given in the third boost. Semi-quantitative analysis of the sera at the peaks of antibody production following each boost showed an approximate fourfold increase in titre per boost. ELISA results (EIA Abbot System, Fig. 1*c*) correlated closely with the WB and demonstrated a dramatic rise of antibody after the third boost. Moreover, RIPA confirmed both the pattern of reactivity with gp160 and gp120 and the quantitative enhancement of antibody titre observed at this stage (data not shown).

An activity that neutralized infectivity of T4 cells by HIV also developed after the primary immunization (Fig. 2*a*, upper panel). Preincubation of virus with D.Z. serum not only inhibits infection of H9 cells by HTLV-III_B, the virus strain used in the immunizations, but also infection with the very divergent strain of HIV-1, HTLV-III_{RF} (Fig. 2*a*, *b*). These viruses were early isolates of HIV-1⁴, and differ from each other by over 1,000 nucleotides¹⁰. The group-specific viral neutralization, low after the primary immunization, became clearly evident after the boost (Fig. 2*a*, middle panel and Fig. 2*b*). Neutralizing activity was shown against both divergent viral strains in experiments where serial dilutions of serum (Fig. 2*a*)¹¹ and virus (Fig. 2*b*) were used. HIV neutralizing activity in sera was Ig-associated, as all activity could be removed by preincubation of the sera with goat anti-human Ig and by absorption with HIV-1 infected but not uninfected H9 cells (data not shown).

Preincubation of virus with D.Z. serum also inhibited HIV-1 induced syncytial formation in the CEM-SS cell line at various times following the boost (data not shown). Inhibition was

marked for both HTLV-III_B and HTLV-III_{RF} strains, but was significantly higher for HTLV-III_B. Inhibition of syncytia induced by still another divergent strain of HIV, HTLV-III_{MN}¹² was also obtained.

Specific cell-mediated immunity (CMI) against HIV-1 was observed after the primary immunization¹, and was markedly enhanced after each boost. This was determined by measuring the proliferation of fresh PBL from D.Z. in response to stimulation with either HTLV-III_B or HTLV-III_{RF} (Fig. 3). The percentage of these cells expressing interleukin-2 (IL-2) receptors was also measured using the Tac monoclonal antibody. As previously reported for the V25 priming¹ the increased number of cells expressing IL-2R paralleled the proliferative index (data not shown). After the third boost a slight proliferative response was also obtained with HIV-2 (Fig. 3). A small amphipathic peptide (amino acids 428–443) of gp120, called T1, has previously been shown to induce CMI response against HIV in mice¹³. The proliferative response of D.Z. PBL to T1 increased markedly following the third boost (J. Berzowski, personal communication).

Cell mediated cytotoxicity (CMC) against HIV-1 infected cells was investigated by the chromium release assay using D.Z. PBL as effector cells and histocompatible (HLA-A3 phenotype) uninfected H9 (control) and HTLV-III_B-infected H9 (experimental) cells as targets. After each boost the D.Z. effector cells transiently lysed H9-HTLV-III_B targets to a far greater extent than uninfected H9 targets, whereas PBL from the non-immunized HLA-A3 control individual did not (Fig. 4). Cytotoxicity against HIV-infected targets decreased markedly in the presence of anti-CD8 monoclonal antibodies (data not shown). The cytotoxicity was not exclusively HLA-restricted, however, as non HLA-identical effector cells from the other immunized individuals also lysed HIV-infected H9 targets to a much greater extent than uninfected H9 target cells. In addition, K562 cells, the classic NK targets, were sensitive to lysis by PBL collected from D.Z. 24 days after the first boost (A. Bensoussan and I. Fossati, personal communication). Finally, antibody-dependent cell-mediated cytotoxicity (ADCC) against CD4-bearing cells

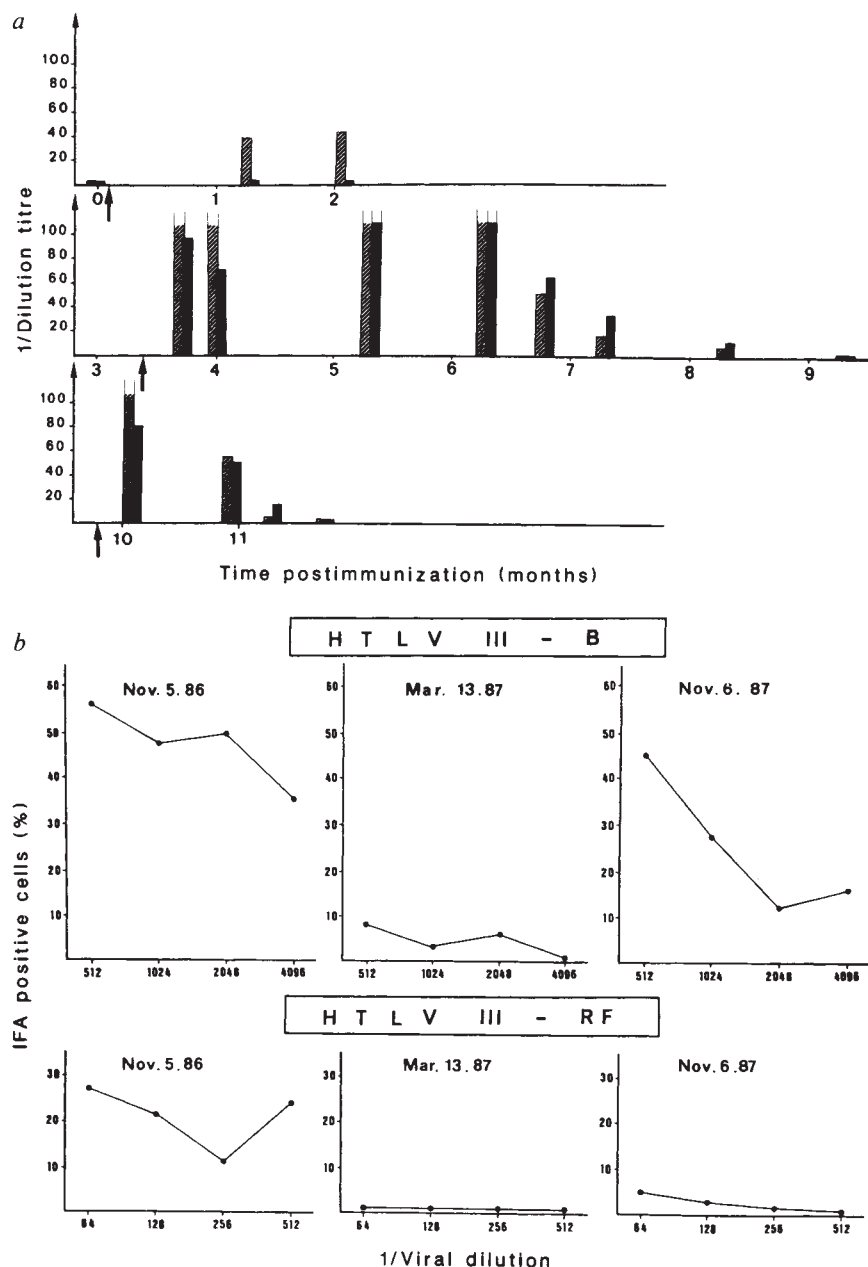


Fig. 2 Serum neutralizing antibodies to HIV, obtained from D.Z. at various periods before and after various stages of immunization and tested for their capacity to inhibit infection of H9 cells by the homologous strain of HIV, HTLV-III_B, and the divergent strain HTLV-III_{RF}. Estimation of virus inhibition of virus was assessed by indirect immunofluorescence (IFA) of infected cultured cells utilizing anti-HIV p24 MAb followed by treatment with goat anti-mouse Ig Ab coupled to fluorescein, as described by Robert-Guroff *et al.*¹¹. *a*, Serum neutralizing titres required to inhibit infection of H9 cells by 50% using a constant viral dilution of HTLV-III_B (hatched bars) and HTLV-III_{RF} (solid bars)¹¹. Arrows indicate times of immunization. *b*, Neutralizing activity against various dilutions of HTLV-III_B (top row) and HTLV-III_{RF} (bottom row). The sera were collected at three different stages: before immunization (5 November 1986), 24 days after the first boost (13 March 1987) and 13 days after the third boost (6 November 1987).

Methods. H9 cells were cultured with suspensions of either HTLV-III_B or HTLV-III_{RF} which had been preincubated with the test serum first for 30' at 4° and then 30' at room temperature. HIV infection was measured by determining the percent of IF positive cells 5 or 6 days later. H9 and H9-infected cells used as controls showed 0 and 70–90% IF positive cells, respectively. The stock HTLV-III_B and HTLV-III_{RF} viral suspensions had reverse transcriptase activities of $5\text{--}10 \times 10^5$ and $1\text{--}2 \times 10^5$ c.p.m. per ml respectively as measured using an assay system previously described¹⁶. Inhibition of syncytia formation as reported by Nara *et al.*¹⁵ was also analysed (data not shown).

(CEM) with bound gp120 was found using normal PBLs as effector cells and sera collected from D.Z. after the immunizations (D. Bolognesi and K. Weinhold, personal communication).

In vivo immunologic tests were performed before the second boost (day 299) and 20 days after the third boost (day 370) by intradermal injection of purified gp160 (10 µg). Immediate allergic hypersensitivity, characterized by a 40 mm diameter erythema with local edema, was observed 1 h after injection and lasted 2 days. In addition, a small (3–4 mm in diameter) red papule appeared at the site of injection after 48 h and lasted 6–7 days in both instances indicating a mild delayed hypersensitivity reaction.

The different treatments used in this immunization protocol did not generate any clinical side effects. Similarly, we found no toxic side effects with previous immunizations of two types of monkeys and chimps by the same approach (data not shown). Moreover, there were no detectable changes in immune parameters (CD4 and CD8 cell counts and IL-2 production from PHA-activated PBL) either after the initial immunization or after subsequent ones.

We suggest that this approach be considered as a prototype candidate vaccine against AIDS. However, whether the above described immune state against HIV confers protection against AIDS has to be established. Indeed, an estimation of the efficacy of a vaccine for protection against AIDS can only be achieved after a large scale clinical trial in a group of volunteers with a high risk of natural infection and followed for an adequate period of time, as has been recently done for typhoid fever vaccination trials in Nepal¹⁴. The protocol used for D.Z. presents logistical problems in any study of a large number of individuals (because of the use of autologous V25 infected cells) and we are presently attempting to mimic the response described here with related but technically more feasible approaches for large scale clinical trials.

We thank Drs V. Kalyanaraman for performing some of the radioimmune precipitation assays and M. Robert-Guroff (NCI, NIH, Bethesda), D. Bolognesi and T. Matthews (Duke University) for helpful suggestions. Initial preparations of recombinant vaccinia virus (V25) were provided by B. Moss (NIAID, NIH, Bethesda, MD) for animal studies and subsequent preparations by G. Beaud (CNRS, Paris).

Fig. 3 Induction of cell-mediated immunity (CMI) against HIV. CMI against HIV was assayed during the course of immunization by estimating the activation of fresh blood mononuclear cells (PBL) with irradiated (2,500 joules) HTLV-III_B and HTLV-III_{RF}. The proliferative index was obtained by the mean result of triplicate samples divided by the control value (no virus). At each time-point, the results shown correspond to the viral dilutions ($1-5 \times 10^{-2}$) which gave the maximum proliferative response (*) D.Z. samples; (○) normal non-immunized individual samples; (→) time of the boost (see text).

Methods. PBL activation was performed as described by Zarling *et al.*¹⁷. In brief, fresh PBL from D.Z. and from a normal non-immunized individual obtained at several times during the course of the immunization of D.Z. were isolated by Ficoll-hypaque and suspended in RPMI supplemented with 10% heat-inactivated normal human serum. Medium (0.1 ml) containing various dilutions ($10^{-1}-10^{-3}$) of UV-irradiated (2500 joules) HIV-1 (strains HTLV-III_B and HTLV-III_{RF}) or HIV-2 (strain NIH-DZ)¹⁸, derived from stock suspensions of virus

($1 \mu\text{g ml}^{-1}$ centrifuged from supernatants of infected H9 cells and exhibiting reverse transcriptase activity of $5-8 \times 10^5$ c.p.m. per ml) was added to 0.1 ml aliquots of 3×10^4 cells. The cells were cultured at 37 °C in a 5% CO₂ incubator. Six days after stimulation ³H-thymidine (1.5 μCi) was added for 18 h, and its incorporation into DNA was measured in a β -counter.

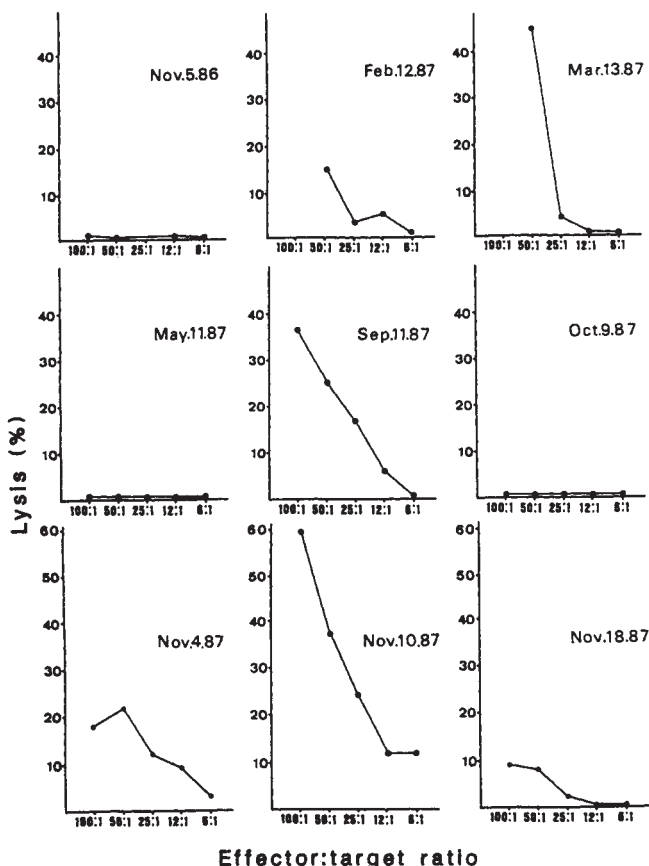
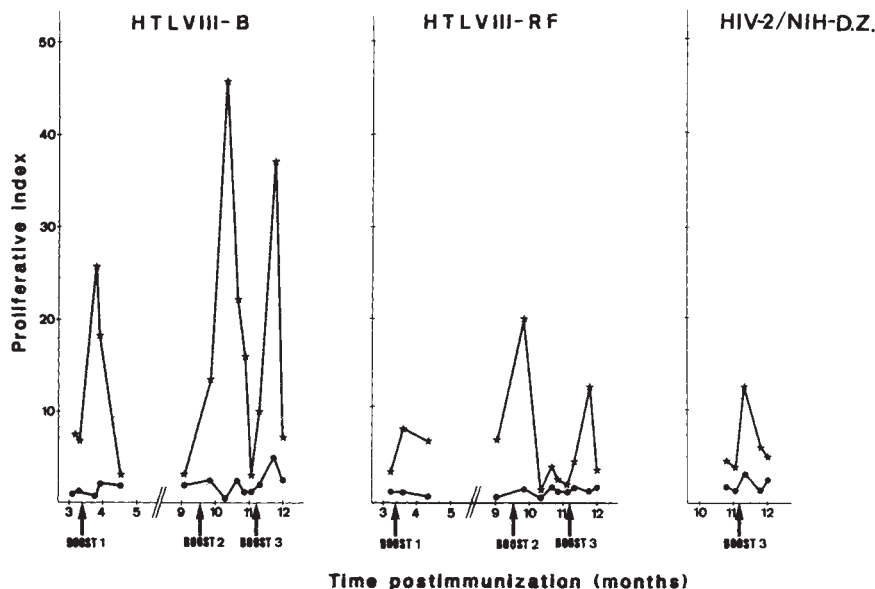


Fig. 4 Cell-mediated cytotoxicity against HIV infected cells. Peripheral blood mononuclear cells (recovered after freezing) and labelled targets (HTLV-III_B infected or uninfected H9 cells) were mixed in duplicates at various effector: target ratios and incubated 5 h at 37 °C for a chromium-release assay. The results are expressed as a percentage of the difference between lysis of HTLV-III_B-infected and uninfected H9 cells and represent the mean of at least two experiments. In these assays lysis of control uninfected H9 cells ranged between 5 and 20% at an effector:target ratio of 100.

Received 28 January; accepted 10 March 1988.

1. Zagury, D. *et al.* *Nature* **326**, 249-250 (1987).
2. Chakrabarti, S. *et al.* *Nature* **320**, 535-537 (1986).
3. Hahn, B. H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 4813-4817 (1985).
4. Popovic, M. *et al.* *Science* **224**, 497-500 (1984).
5. Wong-Staal, F. & Gallo, R. C. *Nature* **317**, 395-403 (1985).
6. Fuerst, T. R., Earl, P. L. & Moss, B. *Molec. cell Biol.* **7**, 2538-2544 (1987).
7. Gallo, R. C. *et al.* *Science* **224**, 500-503 (1984).
8. Zagury, D. *et al.* *Science* **231**, 850-853 (1985).
9. Leonard, R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
10. Starcich, B. R. *et al.* *Cell* **45**, 637-648 (1986).
11. Robert-Guroff, M. *et al.* *Nature* **316**, 72-74 (1986).
12. Wong-Staal, F. *et al.* *Science* **229**, 759-762 (1985).
13. Cease, K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 4249-4253 (1987).
14. Acharya, A. L. *et al.* *N. Engl. J. Med.* **317**, 1101-1109 (1987).
15. Nara, P. L. *et al.* *AIDS Res. hum. Retroviruses* (in the press).
16. Salahuddin, S. Z. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 5530-5534 (1985).
17. Zarling, J. *et al.* *Nature* **323**, 344-346 (1986).
18. Franchini, G. *et al.* *AIDS Res. hum. Retroviruses* **3**, 11-17 (1987).

Transduction of endogenous envelope genes by feline leukaemia virus *in vitro*

Julie Overbaugh, Norbert Riedel*, Edward A. Hoover† & James I. Mullins‡

Department of Cancer Biology, Harvard School of Public Health, 665 Huntington Ave., Boston, Massachusetts 02115, USA

† Department of Pathology, Colorado State University, Fort Collins, Colorado 80523, USA

Feline leukaemia viruses (FeLV) are exogenous retroviruses that can be detected in most cats with leukaemia, aplastic anaemia, myeloproliferative diseases and fatal immunosuppression^{1,2}. FeLV isolates have been divided into three subgroups, based on the viral envelope-determined properties of interference and host range *in vitro*^{3,4}. FeLV-A is present in all natural isolates^{4,5} and is generally minimally pathogenic^{6,7}. FeLV-B is found with FeLV-A in isolates from ~40% of natural infections and in a higher percentage of cats with lymphoma⁵. Following the fundamental observations of

* Present address: Boston University School of Medicine, 80 East Concord Street, Boston, Massachusetts 02118, USA.

‡ To whom correspondence should be addressed.

genetic reassortment of avian retroviruses with endogenous viral genes⁸ and the origination of lymphomagenic viruses during the ontogeny of AKR mice⁹, we show here that transfection of feline cells with FeLV-A DNA results in its recombination with endogenous FeLV-related sequences to produce viruses with the structural and host range properties of FeLV-B. Thus *in vitro* propagation of a retrovirus may result in the generation of variants with very different properties.

We have observed the development of restriction site polymorphisms in proviral DNA present in feline embryo fibroblast (FEF) cells transfected with molecularly cloned FeLV. Using a hybridization probe specific for the long terminal repeat (LTR) of exogenous FeLV (exU3) (ref. 10), we have observed internal viral fragments other than those expected in nine out of ten cell lines transfected with: a minimally pathogenic FeLV from subgroup A (61E, also referred to as F6A (refs 6, 7); a virus chimaera (F3A) (ref. 6) containing the envelope gene of another minimally pathogenic subgroup-A FeLV, and a subgroup-C virus that induces aplastic anaemia (FSC)¹¹. The internal 3.4 kilobase (kb) *Kpn*I fragment of the infecting provirus was detected 2–3 weeks after transfection, and the extra bands were typically detected after 6–10 weeks (Fig. 1). These polymorphic sites were found in the 3' half of the viral genome, but not in the 5' half (compare the *Pst*I/*Hind*III fragment in Fig. 1). In parallel transfections of a dog osteosarcoma-derived cell line (D17) which does not contain endogenous FeLV-related sequences, we did not observe virus restriction-site polymorphisms in cultures maintained for up to six months.

Several lines of evidence suggest that FeLV-B and C arise by recombination. First, FeLV-B and C are found only in association with FeLV-A (ref. 12). Second, FeLV-B-specific sequences have been detected in the DNA of uninfected cats¹³, and these share 99% similarity with regions of the envelope gene of subgroup-B FeLVs (J.I.M. *et al.*, in preparation). Given this evidence, we wanted to determine if the novel viruses that we observed were the result of exogenous FeLV acquiring endogenous FeLV-related (enFeLV) sequences. We therefore analysed viral DNA using a probe that specifically recognizes the envelope gene of FeLV-B and the major family of enFeLV (probe B/S, ref. 13). To discriminate between enFeLV sequences acquired by FeLV and enFeLV sequences at their normal cellular loci, we attempted to transmit virus from FEF cell-culture supernatants to D17 cells. DNA hybridizing to the B/S probe was detected when D17 cells were infected with supernatant from 61E-transfected FEF cells containing polymorphic viral sequences, but not when D17 cells were transfected with 61E DNA without passage through cat cells (Fig. 2). This demonstrated that the new viral genomes contain gp70 sequences derived from enFeLV. Results obtained with 61E were duplicated with FeLV-A isolate F3A, and with FeLV-C isolate FSC, indicating that the capacity for recombination with enFeLV is a general feature of FeLV. We did not observe *Kpn*I site polymorphisms using the exU3 probe in three out of six FEF cell lines transfected with 61E. However, cell-free viral supernatant from two of these cell lines readily established infection in D17 cells, and we detected a comigrating 3.4-kb *Kpn*I viral fragment which specifically hybridized to the B/S probe 14 days after infection (Fig. 2, lanes 2). In the third case, we detected 61E viral DNA only after 10 weeks post-infection of D17 cells, and observed no specific hybridization to the B/S probe, suggesting that this culture contained only 61E with the expected limited ability to infect D17 cells.

Because hybridization studies indicated that the recombinant viruses acquired subgroup B specific sequences, we determined whether their host range was like that of natural isolates of FeLV-B. Like FeLV-B, these recombinant viruses were infectious for dog (D17) cells (which can be infected by FeLV-B and C, but not readily by FeLV-A³) but not guinea pig fibroblast cells (which can be infected by FeLV-C only^{3,11}). The failure to detect viruses with the host range of FeLV-C indicates that sequences

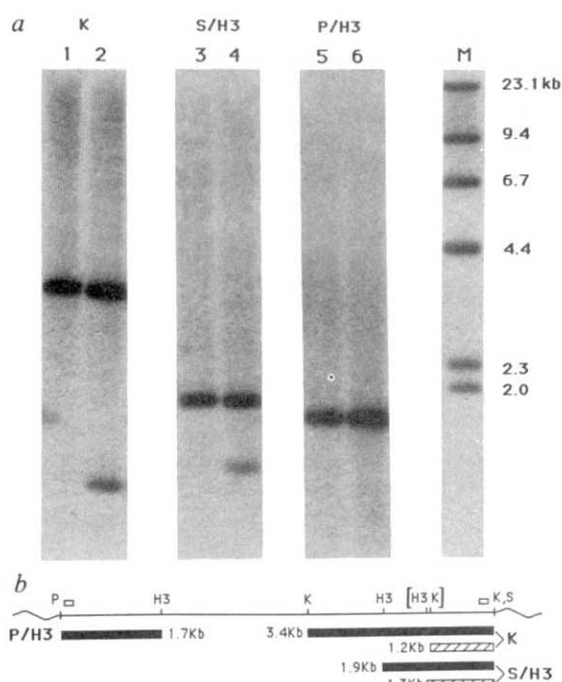


Fig. 1 *a*, Polymorphic viral sequences in feline embryo fibroblast cells transfected with molecularly cloned 61E virus. Blot-hybridization studies of DNA from FEF cells 3 weeks post-transfection with 61E showed only the expected internal viral restriction endonuclease fragments that hybridize to the exU3 probe (3.4-kb *Kpn*I fragment, lane 1; 1.9-kb *Sma*I/*Hind*III fragment, lane 3; 1.7 kb *Pst*I/*Hind*III fragment, lane 5). Restriction site polymorphisms mapping to the envelope region were detected in viral DNA at 10 weeks post-transfection (1.2-kb *Kpn*I fragment, lane 2; 1.3-kb *Sma*I/*Hind*III fragment, lane 4), but not in the *gag* region (lane 6). Size markers are shown in lane M. *b*, A schematic representation of the restriction site map of 61E provirus. The restriction sites shown in brackets correspond to sites not found in the original cloned virus, and only restriction sites relevant to this analysis are shown (H3, *Hind*III; K, *Kpn*I; P, *Pst*I; S, *Sma*I). Cellular sequences flanking integrated provirus are shown by wavy lines. Regions of hybridization with the exU3 probe are shown by open boxes above the map. Below the map the 61E-derived restriction fragments are depicted by filled boxes, whereas the new fragments are shown as striped boxes.

Methods. High-passage FEF cells (line AH927, originated by W. Nelson-Rees and provided by S. Rasheed) were transfected with 1–2 µg plasmid DNA containing the 61E viral genome using the calcium phosphate¹⁶ or electroporation¹⁷ technique. Similar recombinants were observed using both methods. Total genomic DNA was digested, blotted and hybridized with a probe specific for the LTR of exogenous FeLV (exU3), as described¹⁰.

determining FeLV-C are not readily acquired from enFeLV.

To determine which regions of the enFeLV genome had been acquired, we generated restriction maps for several recombinant viruses (Fig. 3). Two main types of recombinants involving 61E were detected, and in three out of six transfections, both types were found in the same cell line (see for example, Fig. 2b, lane 3). Comparison of the predominant recombinant, called EARLE (see legend to Fig. 3), with a representative of enFeLV (enFeLV-5; J.I.M. *et al.*, in preparation) and 61E suggests that it has acquired its entire *gp70* gene from enFeLV. Two species with the EARLE structure but having different 5' recombination sites were detected, one near the 3' end of the polymerase gene (*pol*), and the other further upstream in *pol*. The most frequently observed 3' site for recombination maps to the region encoding

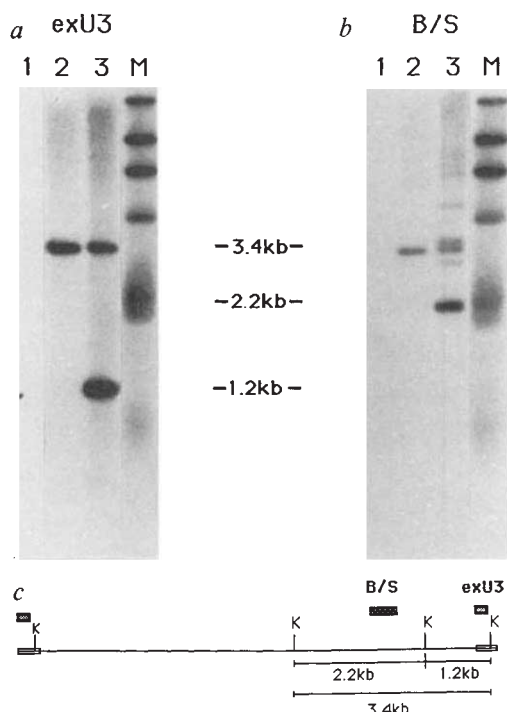


Fig. 2 Blot analysis of proviral DNA from transfected and infected D17 dog osteosarcoma cells. *Kpn*I-digested DNAs were hybridized with probes specific for the exogenous FeLV-LTR (exU3), panel a or the N-terminal half of the *gp70* envelope gene of FeLV-B and enFeLV (B/S probe¹³), panel b. c, Line drawing showing the regions of the FeLV genome from which the probes were derived, and the expected digestion fragments. Proviral DNA was not detected with either probe in D17 cells 8 weeks post-transfection with 61E (lanes denoted 1), but was readily detected with both probes at 14 days after infection with cell-free supernatant from 61E-transfected FEF cells (lanes 2 and 3). The size (in kb) of major bands in lanes 1–3 are indicated. The size markers (lanes M) are the same as in Fig. 1.

Methods. D17-cl dog osteosarcoma cells (J. Riggs) were provided by M. O. Nicolson and transfected as described for Fig. 1. Cells were infected using 1 ml filtered supernatant ($\sim 10^5$ – 10^6 tissue-culture infectious units) from exponentially growing FEF cells transfected ≥ 3 months previously with molecularly cloned 61E. Blot analysis was as described (ref. 10 and Fig. 1), except that filters were washed at 56 °C in $0.2 \times$ SSC (0.03 M sodium chloride, 0.003 M sodium citrate, 0.1% sarkosyl and 0.05% sodium pyrophosphate).

the p15E transmembrane envelope protein. A similar type of recombinant between FeLV-C clone FSC and enFeLV, called SARLE, was identified in each of two FEF cell lines independently transfected with FSC (Fig. 3).

A distinct type of 61E recombinant, called EAGLE, has lost the *Hind*III site found in the central portion of 61E *gp70*, but has not acquired any of the restriction sites typical of the C-terminal encoded portion of *gp70* of the major family of enFeLV (Fig. 3), indicating that the 3' point of recombination occurred within *gp70*. EAGLE has acquired two of the restriction sites (*Bam*HI and *Hind*III, near the 3' end of *pol*) characteristic of enFeLV, indicating that the 5' site of recombination is found upstream of these sites. Some of the EAGLE-type recombinants include a major portion of the endogenous *pol* region (that is, they include the shaded box between the two *Hind*III sites in *pol*).

Genetic reassortment between exogenous retroviruses and between exogenous retroviruses and endogenous viral sequences that result in viruses with new host range properties was originally demonstrated for avian retroviruses¹⁴. This process

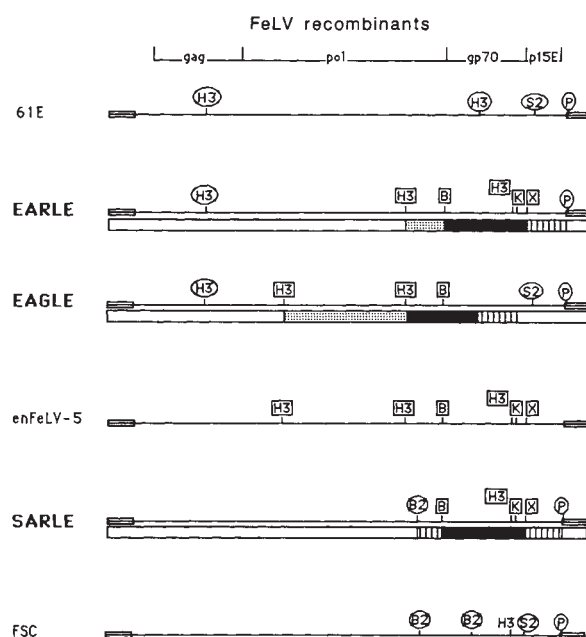


Fig. 3 Restriction site maps of FeLV-recombinants generated in tissue culture. Maps of the main species of recombinant FeLVs were generated by analysis of viral DNA from infected D17 cell lines using exU3, B/S and the entire 61E genome as hybridization probes. Recombinants of 61E were referred to as EARLE and EAGLE (for 61E-associated FRB-like endogenous and 61E-associated FGB-like endogenous, respectively) because their restriction maps were like those of FeLV isolates FRB (the FeLV-B component of the Rickard strain; ref. 18 and J.I.M. *et al.*, in preparation) and FGB (the FeLV-B component of the Gardner-Arnstein strain^{19,20}). Similarly, a FRB-like recombinant of FSC is called SARLE (FSC-associated FRB-like endogenous). Only restriction sites useful in distinguishing the parental and endogenous genomes are shown; sites characteristic of the parental virus are circled and sites unique to a representative of the major family of enFeLV (enFeLV-5; J.I.M. *et al.*, in preparation) are boxed. Below the map, open boxes refer to regions derived from the parental virus, whereas black boxes indicate regions derived from enFeLV. The striped boxes represent the smallest region to which the recombination site could be mapped. Regions of enFeLV found in a minority of the recombinants are shown as stippled boxes. The restriction sites are: B, *Bam*HI; B2, *Bgl*II; H3, *Hind*III; K, *Kpn*I; P, *Pst*I; S2, *Sac*II; X, *Xho*I.

was shown to depend on transcription of the endogenous sequences⁸. We have found that low amounts of enFeLV messenger RNA can be detected in FEF cells (data not shown), allowing the possibility of a similar mechanism for the generation of recombinant FeLVs. The recombinants observed here are also similar in general structure to the tumorigenic mink-cell focusing (MCF) viruses generated from endogenous proviruses during ontogeny in AKR mice⁹, and like the MCF viruses, they have multiple sites of recombination.

The rapid emergence of recombinant viruses as prominent species in cell culture was unexpected, because FeLV-B is not found in most cats carrying FeLV-A (ref. 5). Furthermore in only one of two cats, inoculated with 61E and followed for two years, did we detect evidence of FeLV-B (at a low level), and that could have been present in the inoculum as 61E was propagated in feline cells before inoculation⁷. The higher rate of detection of FeLV-B in FEF cells than in cats could result from possible differences in the immune response to FeLV-A and -B envelope antigens¹⁵, from different amounts of transcription of enFeLV in target cells *in vivo*, or from a restricted host

range of the recombinants in the animal. In any case, our findings indicate that FeLV-B viraemia could be more restricted than assumed at present because natural virus isolates are normally propagated in cat cells before subgroup analysis^{4,5}.

When inoculated into cats, molecularly cloned FeLV-B (FGB) seems to produce a variety of proliferative diseases (including lymphoma, sarcoma and myeloproliferative disease) with various latent periods (E.A.H. *et al.*, in preparation). But it may be difficult to investigate how the formation of FeLV-B and other recombinants relates to the onset of disease. We have recently obtained clones of FeLV directly from the tissues of cats with FeLV-induced fatal immunodeficiency and have shown that this disease is caused by a replication-defective virus which might not have been detected had the virus first been propagated *in vitro*⁷. Our experiments add to the difficulty, here, because we find that new viruses can frequently be created during propagation of cloned virus in tissue culture. We therefore believe that the genesis of complex retroviral diseases, such as feline leukaemia and AIDS, should be studied directly if possible, using procedures such as *in situ* hybridization, immunocytochemistry, and the recovery of virus by molecular cloning directly from the tissues of the infected host.

We thank Dr John Cairns for helpful discussions and James Neil for providing the B/S probe. This work was supported by grants to J.O. from the American Foundation for AIDS Research, and to E.A.H. and J.I.M. from the National Institutes

of Health and the Massachusetts AIDS Research Council. N.R. was a fellow of the Deutsche Forschungsgemeinschaft.

Received 29 December 1987; accepted 1 March 1988.

- Hardy, W. D. Jr in *Feline Leukemia Virus* (eds Hardy, W. D. Jr, Essex, M. & McClelland, A. J.) 3-31 (Elsevier, North-Holland, 1980).
- Hoover, E. A., Rojko, J. L. & Olsen, R. G. in *Feline Leukemia* (ed. Olsen, R. G.) 32-51 (CRC, Boca Raton, 1980).
- Jarrett, O., Laird, H. M. & Hay, D. J. *gen. Virol.* **20**, 169-175 (1973).
- Sarma, P. S. & Log, T. *Virology* **54**, 160-169 (1973).
- Jarrett, O., Hardy, W. D. Jr, Golder, M. C. & Hay, D. *Int. J. Cancer* **21**, 334-337 (1978).
- Donahue, P. R. *et al. J. Virol.* **62**, 722-731 (1988).
- Overbaugh, J., Donahue, P. R., Quackenbush, S. L., Hoover, E. A. & Mullins, J. I. *Science* **239**, 906-910 (1988).
- Weiss, R. A., Mason, W. S. & Vogt, P. K. *Virology* **52**, 535-552 (1973).
- Coffin, J. in *RNA Tumor Viruses* (eds Weiss, R., Teich, N., Varmus, H. & Coffin, J.) 1109-1204 (Cold Spring Harbor Laboratory, New York, 1982).
- Mullins, J. I., Chen, C. S. & Hoover, E. A. *Nature* **319**, 333-336 (1986).
- Riedel, N., Hoover, E. A., Gasper, P. W., Nicolson, M. O. & Mullins, J. I. *J. Virol.* **60**, 242-250 (1986).
- Russell, P. H. & Jarrett, O. *Int. J. Cancer* **21**, 768-778 (1978).
- Stewart, M. A. *et al. J. Virol.* **58**, 825-834 (1986).
- Linial, M. & Blair, D. in *RNA Tumor Viruses* (eds Weiss, R., Teich, N., Varmus, H. & Coffin, J.) 649-783 (Cold Spring Harbor Laboratory, New York, 1982).
- Jarrett, O. & Russell, P. H. *Int. J. Cancer* **21**, 466-472 (1978).
- Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456-467 (1973).
- Potter, H., Weir, L. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7161-7165 (1984).
- Rickard, C. G., Post, J. E., deNoronha, F. & Barry, L. M. *J. natn. Cancer Inst.* **42**, 987-1013 (1969).
- Gardner, M. B. *et al. Nature* **226**, 807-809 (1970).
- Mullins, J. I., Casey, J. W., Nicolson, M. O., Burck, K. B. & Davidson, N. *J. Virol.* **38**, 688-703 (1981).

A highly polymorphic locus very tightly linked to the Huntington's disease gene

John J. Wasmuth, Jeffrey Hewitt*, Barbara Smith, Denis Allard*, Jonathan L. Haines†, Douglas Skarecky, Eric Partlow* & Michael R. Hayden*

Department of Biological Chemistry, University of California, Irvine, California 92717, USA

* Department of Medical Genetics, University of British Columbia, Vancouver V6T 2B5, Canada

† Department of Medical Genetics, Indiana University School of Medicine, Indianapolis, Indiana 46223, USA

The genetic defect in Huntington's disease (HD), an inherited neuropsychiatric disorder of unknown etiology, has not been defined. The discovery of linkage between HD and the DNA marker D4S10(G8) (ref. 1) raised the possibility of isolating the disease gene on the basis of its chromosomal location, in addition to providing a limited presymptomatic test for the late onset disorder. But it has been difficult to isolate other DNA markers nearer to the HD gene, and this has hampered attempts to identify the disease locus and limited the applicability and accuracy of predictive testing. Recently, several new DNA markers from the region of the genome near the HD gene have been isolated using a directed cloning strategy^{2,3}. We describe here the characterization of one of these new markers, D4S95, a highly polymorphic locus which displays no recombination with the HD gene in the families tested. The high degree of polymorphism at this locus and its proximity to the HD gene make it extremely useful for predictive testing and as a new starting point for attempts to clone the disease gene.

Huntington's disease is a neuropsychiatric disorder inherited as an autosomal dominant trait with complete penetrance but delayed onset of symptoms⁴. Nothing is known about the biochemical basis of the disease or the nature of the affected gene. Recent efforts to develop programmes for pre-symptomatic testing and to isolate the HD gene on the basis of its chromosomal location have concentrated on finding genetic markers linked to the disease gene. Based on its linkage to a random DNA

Table 1 Restriction fragment length polymorphisms for D4S95

Restriction enzyme	Allelic fragments (sizes)	Frequency (N)*	PIC†	Heterozygosity (%)
AccI	A1 (7.0)	0.70 (100)	0.33	40
	A2 (1.5)	0.30		
	B1 (4.5)	0.07 (100)		
	B2 (4.4)	0.23		
	B3 (4.3)	0.10		
MboI	B4 (4.2)	0.50	0.67	54
	B5 (4.1)	0.10		
	C1 (1.2)	0.65 (100)		
	C2 (0.7;0.5)	0.35		
PstI	D1 (8.6)	0.91 (100)	0.36	53
	D2 (1.7)	0.09		
TaqI	E1 (2.3)	0.38 (100)	0.36	14
	E2 (1.7)	0.62		

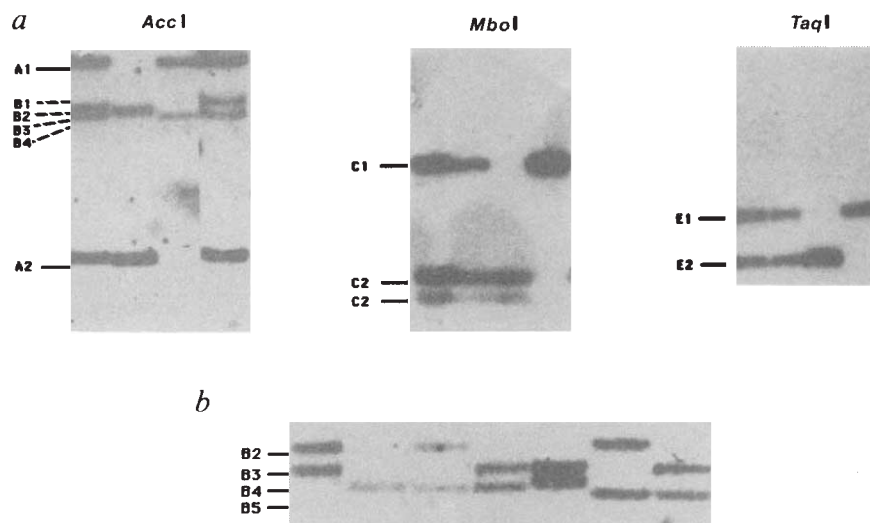
* Number of chromosomes examined.

† Polymorphism information content¹⁴.

marker D4S10 (G8), that is 4 centimorgans from the disease locus and located in 4p16.2-4p16.3, the HD gene has been mapped to the distal portion of the short arm of chromosome 4 (4p) (refs. 3, 5-8). More recent evidence⁹ that the HD gene is between D4S10 and the telomere of 4p indicates the direction from the DNA marker from which to approach the disease gene. D4S62 is another DNA marker 2cM centromeric to D4S10 (refs. 9, 10), and both have been used for predictive testing¹¹, but the significant recombination distance (4-6 cM) of these markers from the disease gene limits their accuracy for testing and their usefulness as starting points for cloning the HD gene. Thus, new DNA markers closer to the disease gene are needed for both diagnostic and research purposes.

We have previously described a cloning strategy to isolate new DNA markers likely to be close to the HD gene^{2,3}. D4S95 is one of these, the closest physically to D4S10 (ref. 3), and proved to be highly polymorphic. As summarized in Table 1 and shown in Fig. 1, the 1-kilobase (kb) probe, BS674E-D defining the genetic locus D4S95, detects five separate poly-

Fig. 1 Allelic restriction fragments at the D4S95 locus. *a*, DNA (2 µg) from unrelated individuals was digested with the indicated restriction endonucleases and then analysed by Southern blot hybridization using the 1.0 kb single-copy DNA fragment, BS674E-D (ref. 3) as the probe. The rare RFLP revealed by *Pst*I (Table 1) is not shown. *b*, DNA from unrelated individuals was digested with *Acc*I and electrophoresed using conditions to enhance the resolution of DNA fragments in the size range (4–5 kb). The multiple alleles at the polymorphic B site were analysed by Southern blot hybridization as in *a*. Allelic fragments B2–B5 are clearly distinguishable. The fifth and largest allelic B1 fragment is not seen in any of these individuals but is apparent in the last lane in the left panel in *a*.



morphic sites using four different restriction endonucleases. *Acc*I reveals two different polymorphic sites, one of which (site A) is a two-allele site polymorphism. The second restriction fragment length polymorphism (RFLP) detected by *Acc*I (site B) is a multi-allele system in which at least five different allelic fragments have been observed (Fig. 1a, top left; Fig. 1b) analogous to the VNTR (variable number of tandem repeats) markers¹². None of the polymorphisms is in complete linkage disequilibrium and all but the *Acc*I site A and the *Mbo*I RFLPs appear to be in complete equilibrium. Among 50 unrelated individuals of Caucasian descent tested with D4S95, 90% were heterozygous for one or more of the RFLPs. Thus, D4S95 provides an informative marker for predictive testing and is especially useful as all of the RFLPs are revealed with a single DNA probe.

Southern blot analysis of a panel of six somatic-cell hybrids containing chromosomes 4 with overlapping deletions of different portions of 4p16 has shown that D4S10 and D4S95 are in the same small physical region, namely 4p16.1–4p16.3 (ref. 3). An important question not soluble by physical mapping, was whether D4S95 is in the direction of the *HD* gene from D4S10. To determine the orientation of D4S95 relative to D4S10 and the *HD* gene, the segregation of D4S95 was examined in seven *HD* families in which individuals with crossovers between D4S10 and *HD* had previously been identified. Each of the recombinants in question has the disease but inherited the non-disease-linked allele for D4S10. The haplotypes of members of two of the families with recombinants between *HD* and D4S10 and typed with D4S95 are shown in Fig. 2. In both of the recombinants shown, D4S95 was inherited with the *HD* gene, rather than with D4S10. In fact, D4S95 did not recombine with the *HD* gene in any of the seven recombination events between D4S10 and *HD* we examined. Thus D4S95 is in the direction of the *HD* gene from D4S10.

The segregation of D4S95 was examined in several other *HD* and reference families typed with D4S62 and D4S10 (ref. 10). Figure 3 shows the results of multipoint linkage analysis of the three DNA markers and the *HD* gene. The peak Lod score for linkage between D4S95 and *HD* is 9.16 and occurs at a recombination fraction of 0.00 (1-Lod confidence interval 0.00–0.04), indicating that there was no recombination between the two loci. The peak Lod score for linkage between D4S10 and D4S95 is 8.16 at a recombination fraction of 0.06 (1-Lod confidence interval 0.02–0.10). These data, together with the lack of recombination between D4S95 and *HD* in seven recombinants between D4S10 and *HD*, indicate that D4S95 is very close to the *HD* gene, even though there is apparently no linkage disequilibrium between the two loci. By combining the results of

the multipoint linkage data and physical mapping studies^{9,10}, the order and orientation of the four loci is established as: cen-D4S62-D4S10-[D4S95;*HD*]-4pter. In the absence of recombination between D4S95 and *HD*, it is not possible to determine whether D4S95 is on the same or opposite side of the *HD* gene relative to D4S10. If D4S95 is proximal to the *HD* gene, it is closer to it than D4S10, but if D4S95 is distal to

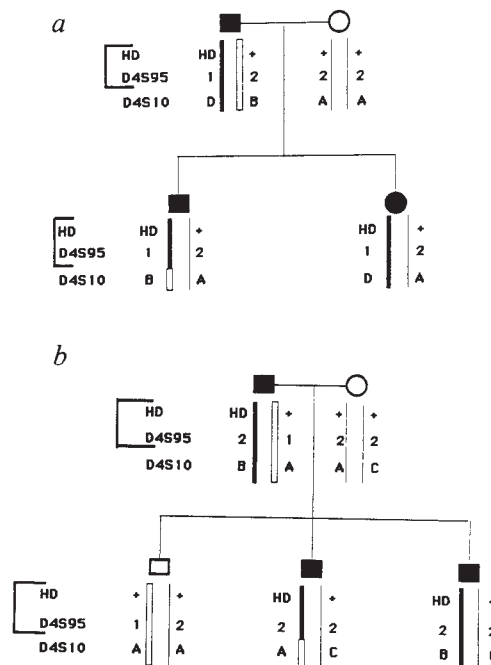


Fig. 2 Segregation of D4S95 in *HD* families with recombinants between D4S10 and *HD*. Open symbols are unaffected individuals, closed symbols are individuals diagnosed as having Huntington's disease. Individual haplotypes were verified by typing other close relatives in extended pedigrees and assuming a minimum number of crossovers between the closely linked loci. The letters listed for D4S10 represent haplotypes (A–D) for the two polymorphic *Hin*-dIII sites revealed by the pK082 subclone¹⁵. The D haplotype for D4S10 is associated with the disease gene for the family shown in *a*, the B haplotype for D4S10 travels with the disease gene for the family shown in *b*. The first sib in family *a* and the second sib in family *b* each inherited the non-disease-linked haplotype for D4S10, yet they both have Huntington's disease. In none of the sibs in either family are there any obligate crossovers between D4S95 and *HD*.

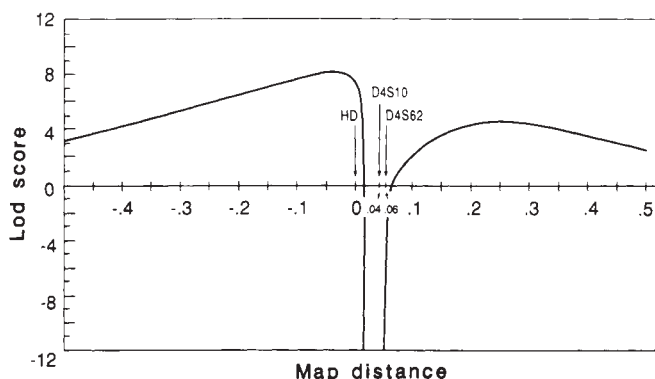


Fig. 3 Multipoint Linkage analysis of *HD*, D4S95, D4S10, and D4S62. The program LINKMAP¹⁶ from the Linkage package (version 3.5) was used to calculate Lod scores for various potential locations of D4S95 relative to previously established positions of *HD*, D4S10 and D4S62. The location of *HD* was arbitrarily set at 0. Lod scores at each map location were approximated in LINKMAP by subtracting the $\log_{10}$ of the likelihood for a distant map position from the $\log_{10}$ of the likelihood at the given location. The peak Lod score between D4S95 and *HD* is 9.16 at a recombination fraction (θ) of 0.00. The favoured location of D4S95 is telomeric to D4S10 by odds of >35,000:1 and the odds are 10:1 that D4S95 is closer to the *HD* gene than D4S10.

the *HD* gene, it may or may not be closer than D4S10 and would represent the first marker flanking the *HD* gene. In either case, the proximity of this highly polymorphic locus to the *HD* gene makes it an accurate and informative marker for predictive testing, as well as providing a new starting point close to the *HD* gene from which to attempt to clone the disease locus.

Recently, another new polymorphic locus, D4S43, has been identified¹³ which is also in the direction of the *HD* gene from D4S10. Although recent pulsed-field gel electrophoresis experiments have shown that D4S95 and D4S43 are separated by at least 100 kb (J. J. W., H. Lehrach and J. Gusella, unpublished results), the relative orientation of the two to each other and to the *HD* gene is unknown. As well as being a more informative genetic marker than D4S43, D4S95 will be more useful for predictive testing because all the RFLPs at the locus are detected with a single DNA probe. In contrast, seven different DNA fragments from the D4S43 locus are required to achieve an overall polymorphism information content¹⁴ of 0.70¹³.

We thank Drs James Gusella, Susan Folstein, Haig Kazaian, Sandra Youngman and John Phillips for providing DNA from families. This work was supported in part by the NIH, The Hereditary Disease Foundation, the MRC of Canada, NHRDP, The Woodward Foundation of Vancouver and The Huntington's Society of Canada.

Received 1 February; accepted 4 March 1988.

- Gusella, J. F. *et al.* *Nature* **306**, 234–238 (1983).
- Wasmuth, J. J., Carlock, L. R., Smith, B. & Immken, L. L. *Am. J. hum. Genet.* **39**, 397–403 (1986).
- Smith, B. *et al.* *Am. J. hum. Genet.* **42**, 335–344 (1988).
- Hayden, M. R. *Huntington's chorea* (Springer, New York, 1981).
- Gusella, J. F. *et al.* *Nature* **318**, 75–78 (1985).
- Magenis, R. E. *et al.* *Am. J. hum. Genet.* **39**, 383–391 (1986).
- Wang, H. S., Greenberg, C. R., Hewitt, J., Kalousek, D. & Hayden, M. R. *Am. J. hum. Genet.* **39**, 392–396 (1986).
- Landegent, J. E. *et al.* *Hum. Genet.* **73**, 354–357 (1986).
- Gilliam, T. C. *et al.* *Cell* **50**, 565–571 (1987).
- Hayden, M. R. *et al.* *Am. J. hum. Genet.* **42**, 125–131 (1988).
- Hayden, M. R. *et al.* *Lancet* **1284**–1286 (1987).
- Nakamura, Y. *et al.* *Science* **235**, 1616–1622 (1987).
- Gilliam, T. C. *et al.* *Science* **238**, 950–952 (1987).
- Botstein, D., White, R. L., Skolnick, M. & Davis, R. *Am. J. hum. Genet.* **32**, 314–331 (1980).
- Gusella, J. F. *et al.* *Science* **225**, 1320–1326 (1984).
- Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3443–3446 (1984).

Nuclear protein with sequence homology to translation initiation factor eIF-4A

M. J. Ford*, I. A. Anton* & D. P. Lane†

Imperial Cancer Research Fund Clare Hall Laboratories, South Mimms, Potter's Bar, Herts EN6 3LD, UK

The p68 protein is a highly conserved nuclear antigen that is thought to be important in the regulation of cell growth and division. It is found in dividing cells of all mammals and amphibians tested, but not in quiescent cells. The protein shows a distinct granular distribution in the nucleus and is induced within four hours of serum stimulation of quiescent mouse fibroblasts. The p68 protein was first identified because of its specific immunological cross-reaction with the DNA tumour virus nuclear oncogene SV40 large T, detected with the anti-SV40 large T monoclonal antibody DL3C4 (ref. 1), now renamed PAb204 (ref. 2). Sequencing of human complementary DNA coding for the growth-regulated p68 nuclear protein has revealed the molecular basis for its cross-reaction with SV40 large T antigen and its extensive homology with the translation initiation factor eIF-4A. The sequence similarity between p68 and eIF-4A is interesting because eIF-4A acts as an ATP-dependent RNA helicase^{3,4} and T antigen is an ATP-dependent DNA helicase⁵. We suggest that p68 could be a DNA or RNA helicase in the cell nucleus which is involved in replication, transcription or RNA processing and is required for cell growth.

The sequence of the p68 protein was determined from a number of independent cDNA clones obtained from two different cDNA libraries. Initially a λ gt11 library made from human hepatoma messenger RNA (ref. 6) was screened using the PAb204 antibody and an immunoperoxidase detection system. The high-density screening procedure used had previously been optimized by constructing a test phage that encoded amino acids 272–708 of SV40 T antigen and expressed the PAb 204 epitope (refs 7, 8, M.J.F. and D.P.L., in preparation). Of the eight million plaques screened, only four gave a positive signal. Restriction analysis and DNA sequencing showed that two of these clones were identical and all three unique clones contained common segments. The inserts from these clones were transferred to *Eco*RI-cleaved pUR290 (ref. 9) to confirm the expression of the epitope by these fragments. The shortest clone (clone λ 1) contained only 150 base pairs (bp) of cDNA, and yet reacted strongly with the antibody. The sequence of the insert was determined by double-stranded sequencing from the pUR plasmid and was confirmed using a conventional template after transfer of the insert to M13. The sequence contained an open reading frame in the correct frame for expression in λ gt11 and pUR. There was a short region of limited amino acid sequence homology (Fig. 1) between this insert and the 64 amino acids of T that contain the PAb204 epitope^{7,8}. This region of homology was responsible for the cross-reaction, as confirmed by the positive reaction of the PAb204 antibody with synthetic peptides corresponding to both the p68 and T antigen sequences (Fig. 1). A more detailed analysis of the immunochemical properties of this highly specific cross-reaction will be published elsewhere (M.J.F. and D.P.L., in preparation).

Restriction mapping of the two remaining λ gt11 clones (λ 2, λ 3) revealed single *Eco*RI sites in the cDNA inserts. This raised the possibility that one or both of these bacteriophage could contain two non-contiguous cDNA inserts as *Eco*RI linkers are used in making λ gt11 cDNA libraries. To investigate this possibility, a second human cDNA library that had been prepared in a plasmid vector without using such linkers^{10,11}, was screened

* Present addresses: Wellcome Biotech, Beckenham, Kent, UK (M.J.F.); Beatson Institute, Bearsden, Glasgow, UK (I.A.A.).

† To whom correspondence should be addressed.

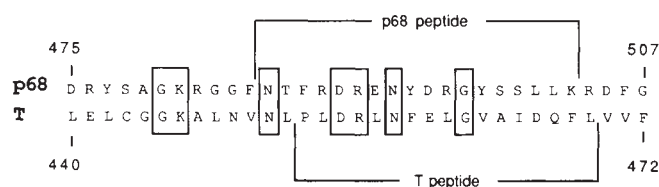


Fig. 1 Cross-reaction between p68 and SV40 large T. The sequence of the 50 amino acids encoded by the λ 1 p68 clone were compared with the sequence of the 64 amino acids of SV40 large T, previously shown to display the epitope for the PAb 204 antibody. The best line-up between the two sequences is shown. On the basis of this alignment a series of synthetic peptides corresponding to sections of the p68 and SV40 large T sequence were synthesized and tested in an enzyme-linked immunosorbent assay for their reaction with the PAb 204 antibody. The shortest p68 and SV40 large T peptides giving a positive reaction are marked.

by hybridization with a clone λ 1 insert probe. The frequency of positive clones in this screen was much higher. One clone containing an insert of 1,300 bp matched λ 3 in sequence on both sides of the internal *Eco*RI site (and throughout), evidence of the contiguous nature of this λ clone. The λ 2 clone in contrast contained a 3' *Eco*RI fragment which overlapped with the λ 3 and plasmid clone, but the sequence of the 5' *Eco*RI fragment was unrelated to that of the other clones. Thus λ 2 may contain two distinct cDNA inserts. The overlapping cDNA clones and selected restriction fragments were subcloned into M13 and sequenced using the dideoxy method. Specific oligonucleotide primers were synthesized to help generate the final sequence. The DNA sequence gave a contiguous sequence of 2,088 bp. The first 1,782 bp contain a continuous open reading frame capable of coding for a protein of 594 amino acids with a predicted relative molecular mass (M_r) of 66,922 (66.9K). Accurate sizing of p68 by SDS-PAGE yields an M_r of 67K, so the deduced amino-acid sequence is probably almost complete. After the stop codon, suitable signals for polyadenylation and a poly(A) tract are found at nucleotides 2,065 and 2,088 respectively.

When we searched the Swiss Prot data base using the deduced amino-acid sequence of p68, we found a remarkable homology to the deduced amino-acid sequence of the murine translation initiation factor eIF-4A¹² (ref. 12). Neither the p68 nor the eIF-4A amino-acid sequences scored significant homology with any other sequence in this data base. Alignment of the two amino-acid sequences demonstrates that the homology extends through the entire eIF-4A sequence, with p68 sequences extending 58 residues out from the N-terminus of eIF-4A and 139 residues from the C-terminus (Fig. 2). Good alignment requires only a single residue gap in the p68 sequence and three small gaps in the eIF-4A sequence (one of four residues, one of three residues and one of a single residue). This alignment identifies 123 identical residues between eIF-4A and p68 out of the 390

p68	F G G S R A G P L S G K K F G N P G E K L V K K K W N L D E	30
eIF-4A	L P K F E K N F Y Q E H P D L A R R T A Q E V E T Y R R S K M E	60
p68	E I T V R G H N C P K P V L N F Y E A N F P A N V M D V I A	90
eIF-4A	P E G V I E S N W N E I V D S E F D D M N L S E S L L R G I Y	
p68	R Q N F T E P T A I Q A Q G W P V A L S G L D M V G V A Q T	120
eIF-4A	A Y G F E K P S A I Q Q R A I L P C I N G Y D V I A Q A Q T	
p68	G S G K T L S Y L L P A I V H T N H Q P F T E R G D G P I C	150
eIF-4A	G T G K T A T F A I S I L Q Q I - - - E L D L - K A T Q A	
p68	L V L A P T R E L A Q Q V I Q Q V A - A E Y C R A C R L K S T	179
eIF-4A	L V L A P T R E L A Q Q I Q K V V M A L G D Y M G A S C H A	
p68	C I Y G G A P K G P Q I R D L E R G V E I C I A T F P G R L I	209
eIF-4A	C T G T N V R A E V Q K L Q M E A P H I V G T P G R V F	
p68	D F L E C G K T N L R R T T Y L V L D E A D R M L D M G F E	239
eIF-4A	D M L N R R Y L S P K Y I K M F L D E A D R M L S R G F K	
p68	F L I R K T V D Q I R P D R Q T L M W S A T W F K E V R Q L	269
eIF-4A	D Q I Y D I F Q K L N S N T Q V V L L S A T M P S D V L E V	
p68	A E D F L K K D Y T H I N I G A L E I L S A N H N I L Q I V D V	299
eIF-4A	T K K F M R D P I R I T L V K K E E I T L E G I R Q F Y I N V	
p68	C H D V E K D E K L I R L M E E I M S E K E N K T I V F V E	329
eIF-4A	E R E E W K L D T L C D L Y E - - - T L T I T Q A V I F I N	
p68	T K R R C D E L T R K M R R D G W P A M G I H G D K S Q Q E	359
eIF-4A	T R R K C D W L T R E K M H A R D T V S A M H G D M D Q R E	
p68	R D V W L N E F K H G K A P I L T A T D V A S R G L D V E D	389
eIF-4A	R D V I M R E F R S G S R V L I T T D L L A R G I D V Q D	
p68	V K F V I N Y D Y F N S S B D Y I H R I G R T A R S T K T G	419
eIF-4A	V S L V I N Y D L P T N R B N Y I H R I G R G G R F G R K G	
p68	T A Y T F F T P N N I K Q V S D L I S V L R E A N Q A I N P	449
eIF-4A	V A I N M V T E E D K R T L R D I E T F Y N T S I E E M P L	
p68	K L L Q V E D R G S G R S R G R G G M K D D R D R Y S A	479
eIF-4A	N V A D L I	
p68	G K R G G F N T F R D R E N Y D R G Y S S L L K R D F G A K	509
eIF-4A		
p68	T Q N G V Y S A A N Y T N G S F G S N F V S A G I Q T S F R	539
eIF-4A		
p68	T G N P T G T Y Q N G Y D S T Q Q Y G S N V P N M H N G M N	569
eIF-4A		
p68	Q Q A Y A Y P A T A A A P M I G Y P M P T G Y S Q	594
eIF-4A		

Fig. 2 Amino-acid sequence homology between human p68 and murine eIF-4A. The deduced sequence of p68 and eIF-4A were aligned. The 123 identical amino acids are boxed and the epitope recognized by the PAb 204 antibody, as determined by studies with synthetic peptides, is heavily underlined (amino-acid positions 486-503). The 17 residues of the type-A consensus¹³ are underlined (positions 116-132), as are the six residues of the modified type-B motif (positions 224-229; I.A.A., M.J.F. and D.P.L., in preparation).

residues of total eIF-4A sequence (32% identity). The regions of identity are spread throughout the sequence and include some long blocks of absolute identity. Further analysis of the two sequences identified 53 very conservative amino-acid substitutions in addition to the 123 identical residues (45% identical, or highly conserved, residues). Local secondary structure predictions suggest a very similar form for both proteins. p68 contains 193 residues, predicted to be in a helical conformation, most of them in a block extending from residue 228 to 390. We also predict that most of this region of eIF-4A is helical (M.J.F.,

Fig. 3 The type-A consensus in p68, eIF-4A and some known DNA helicases. Amino-acid sequences around the type-A consensus are compared in p68, eIF-4A and four prokaryotic DNA helicases (*dna* B, (ref. 14); *rec* B, (ref. 15); *rho*, (ref. 16); *uvr* D, (ref. 17). The presence of alanine rather than glycine at position three of the consensus is noted. The consensus for proteins using adenine nucleotide is: position 1, always uncharged; 3, A or G; 6, usually G; 7, always uncharged (usually M, S, V, T or A); 8, always G; 9, always K; 10, T or S; 13, always uncharged; 17, uncharged, usually hydrophobic (ref. 13; and I.A.A., M.J.F. and D.P.L., in preparation).

		1	3	6	7	8	9	10	13	17	
		x	x	A	x	x	G	x	G	K	T
		x	x	G	x	G	K	T	x	x	x
p68		P	T	A	I	Q	A	Q	G	W	P
eIF-4A		P	S	A	I	Q	Q	R	A	I	L
<i>dna</i> B		G	Y	D	D	L	N	K	K	T	A
<i>rec</i> B		S	D	V	A	E	T	L	D	P	L
<i>rho</i>		L	T	A	R	V	L	D	L	A	S
<i>uvr</i> D		D	S	L	N	D	K	Q	R	E	A

I.A.A. and D.P.L., in preparation). A comparison of the nucleic-acid sequence encoding eIF-4A with that encoding p68 gave 521 matched bases out of 1,168 (44%).

The polypeptide eIF-4A has an M_r of 46K and can melt or unwind mRNA in the presence of ATP (refs 3,4). The protein binds ATP and has been isolated as a subunit of the eIF-4F cap-binding complex, as well as in the free form. The sequences of eIF-4A and p68 contain the type-A motif¹³ found in many proteins using ATP. In Fig. 3, we compare this sequence feature in p68, eIF-4A and a number of prokaryotic helicases. It is interesting that at position 3 of the motif, all these proteins have an Ala rather than the Gly residue found in most proteins displaying the consensus. The type-B motif¹³ in p68 and eIF-4A is also contained in a modified form (M.J.F., I.A.A. and D.P.L., in preparation). Preliminary studies on immunoaffinity-purified p68 from HeLa cells indicate that the protein has an associated ATPase and DNA helicase activity. DNA helicases are necessary for DNA replication and repair processes, and could also be involved in transcription complexes. Further work should help to clarify the relationship between this growth-regulated nuclear protein and eIF-4A, and may lead to the discovery of other members of a family of eukaryotic proteins involved in melting duplex polynucleotides. The cross-reaction between the eukaryotic viral DNA helicase, SV40 large T, and p68 suggests that the common epitope is involved in a common function. This remains unresolved, however, as the epitope on p68 lies on the C-terminal side of the eIF-4A homology. The PAb 204 antibody is able to block the ATPase, helicase and DNA replicative

functions of T. Here the epitope lies close to the type-A motif in the primary structure⁷. The epitope is remote from the motif, in p68, at least in the primary structure, and experiments are in progress to determine the effect of PAb 204-binding on p68 function. It seems intrinsically unlikely that the cross-reaction is fortuitous, given the apparent functional similarity between p68 and large T. It is tempting to speculate that p68 may provide a cellular helicase that is mimicked and supplanted by the viral large T helicase to uncouple viral replication from host control.

We thank Ian Goldsmith for oligonucleotide synthesis, Johnathan Rothbard for peptide synthesis and for discussions of epitope location, and Michelle Ginsburg for assistance with the computer operations.

Received 26 January; accepted 7 March 1988.

1. Lane, D. P. & Hoeffler, W. K. *Nature* **288**, 167-170 (1980).
2. Crawford, L. V. & Harlow, E. *J. Virol.* **41**, 709 (1982).
3. Ray, B. K. *et al. J. biol. Chem.* **260**, 7651-7658 (1985).
4. Abramson, R. D. *et al. J. biol. Chem.* **262**, 3826-3832 (1987).
5. Stahl, H., Droge, P. & Knippers, R. *EMBO J.* **5**, 1939-1944 (1986).
6. deWet, J. R. *et al. DNA* **3**, 437-447 (1984).
7. Mole, S. E., Gannon J. V., Ford, M. J. & Lane, D. P. *Phil. Trans. R. Soc. B* **317**, 455-469 (1987).
8. Mole, S. E. & Lane, D. P. *J. Virol.* **54**, 703-710 (1985).
9. Ruther, U. & Muller-Hill, B. *EMBO J.* **2**, 1791-1794 (1983).
10. Okayama, H. & Berg, P. *Molec. cell. Biol.* **3**, 280-289 (1983).
11. Okayama, H. & Berg, P. *Molec. cell. Biol.* **5**, 1136-1142 (1985).
12. Nielsen, P. J., McMaster, G. K. & Traschel, H. *Nucleic Acids Res.* **13**, 6867-6880 (1985).
13. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. *EMBO J.* **1**, 945-951 (1982).
14. Nakayama, N., Arai, N., Bond, M. W., Kaziro, Y. & Arai, K. *J. biol. Chem.* **259**, 97-101 (1984).
15. Finch, P. W. *et al. Nucleic Acid Res.* **14**, 8573-8582 (1986).
16. Pinkham, J. L. & Platt, T. *Nucleic Acid Res.* **11**, 3531-3545 (1983).
17. Finch, P. W. & Emmerson, P. T. *Nucleic Acid Res.* **12**, 5789-5799 (1984).

The binding site for C1q on IgG

Alexander R. Duncan* & Greg Winter

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

In humoral defence, pathogens are cleared by antibodies acting as adaptor molecules: they bind to antigen and trigger clearance mechanisms such as phagocytosis, antibody-dependent cell-mediated cytotoxicity and complement lysis¹. The first step in the complement cascade is the binding of C1q to the antibody. There are six heads on C1q, connected by collagen-like stems to a central stalk², and the isolated heads bind to the Fc portion of antibody rather weakly, with an affinity of 100 μ M (ref. 3). Binding of antibody to multiple epitopes on an antigenic surface, aggregates the antibody and this facilitates the binding of several C1q heads⁴, leading to an enhanced affinity of about 10 nM (ref. 1). Within the Fc portion of the antibody, C1q binds to the CH₂ domain^{5,6}. The interaction is sensitive to ionic strength⁷, and appears to be highly conserved throughout evolution as C1q reacts with IgG from different species (for example see ref. 8). By systematically altering surface residues in the mouse IgG2b isotype, we have localized the binding site for C1q to three side chains, Glu 318, Lys 320 and Lys 322. These residues are relatively conserved in other antibody isotypes, and a peptide mimic of this sequence is able to inhibit complement lysis. We propose that this sequence motif forms a common core in the interactions of IgG and C1q.

Several sites of C1q binding have been proposed, based on chemical modification studies, inhibition of complement lysis by synthetic peptides and oligosaccharides, and sequence comparisons of antibody isotypes¹. Taken together, however, these sites cover most of the solvent accessible surface of the CH₂ domain (Fig. 1a). Recently the binding site for tRNA on the tyrosyl tRNA synthetase was localized by 'surface scanning'—

systematically removing side chains and scoring the mutant enzymes for altered binding of tRNA⁹. We have used this technique to localize the C1q binding site in antibodies, focusing on conserved residues capable of polar interactions.

We first constructed and expressed a series of mutant antibodies (Fig. 2), starting from a chimaeric antibody with a heavy chain of mouse IgG2b isotype¹⁰ that binds to the hapten NIP (3-iodo-4-hydroxy-5-nitrophenylacetate). The 'parent' antibody binds to C1q and participates in complement lysis¹. Mutant antibodies were therefore scored for their ability to mediate complement-dependent lysis of sheep red blood cells coated with the hapten NIP. Most of the mutants were lytic, except for Asn 297 $\rightarrow$ Ala, Glu 318 $\rightarrow$ Ala; Lys 320 $\rightarrow$ Ala and Lys 322 $\rightarrow$ Ala. Asparagine 297 is located at the top of the CH₂ domain and is the carbohydrate attachment site; Glu 318-Lys 322 are located close together on a strand of β sheet (Fig. 1b).

These four non-lytic mutants were then scored for binding to C1q. Antibody was aggregated by binding to a solid support (NIP-Affigel) and the affinity determined by measuring the binding of ¹²⁵I-labelled C1q (Table 1). The affinity of the mouse IgG2b antibody was 11 nM, as expected¹. However, the affinity of the Asn 297 $\rightarrow$ Ala mutant for C1q was threefold weaker than the wild type. Removing carbohydrate from antibody by glycosidases¹¹ or by treating the cells with the glycosylation inhibitor tunicamycin, destroys the ability of the mouse IgG2b isotype to fix complement^{12,13}, and reduces the affinity for C1q about threefold¹⁴. Presumably the effect of the Asn 297 $\rightarrow$ Ala mutation on lysis and C1q binding is due to the lack of carbohydrate. The affinities of the three other mutants, Glu 318 $\rightarrow$ Ala, Lys 320 $\rightarrow$ Ala and Lys 322 $\rightarrow$ Ala for C1q were at least 30-fold lower than the wild type. Thus C1q appears to interact with the charged side chains of residues 318, 320 and 322, but not with the adjacent hydrophilic residues 274, 283, 317, 326, 333, 335, 337 and 340 (Fig. 1b). The region 318-337 is one of the segments of the CH₂ domain which has been proposed as a C1q binding site⁷ (Fig. 1b).

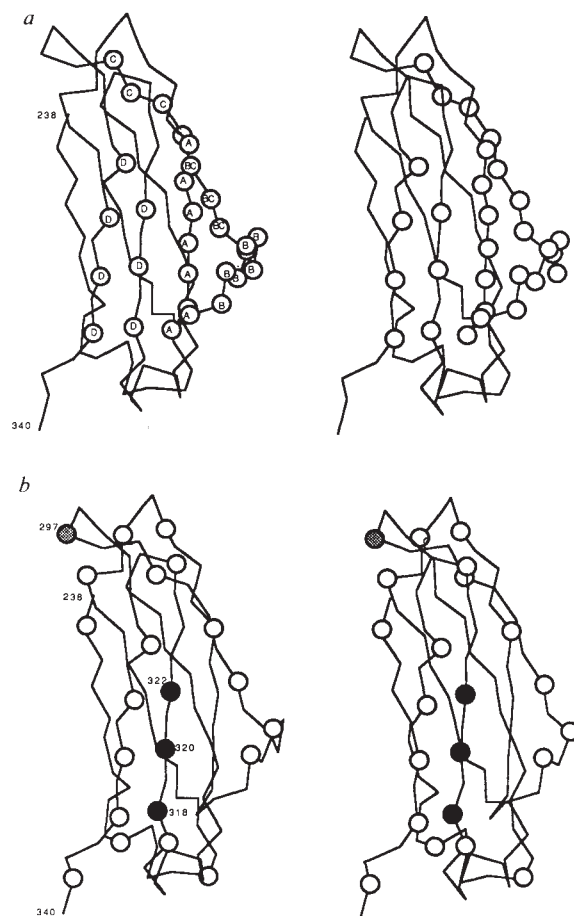
Residues Glu 318, Lys 320 and Lys 322 are conserved in all the human IgGs, rat IgG2b and IgG2c, mouse IgG2a, IgG2b and IgG3, guinea pig IgG1 and rabbit IgG. However, in three isotypes there are changes: mouse IgG1 (Arg322), rat IgG1

* Present address: Department of Biology Q-063, University of California, San Diego, California, USA.

Fig. 1 (right) α -Carbon trace of the CH₂ domain from human Fc²³. **a**, Residues which have been proposed as C1q binding sites. A, residues 274–281¹⁹; B, residues 282–292²⁰; C, residues 290–295²⁴; D, residues 318, 320, 322, 331, 333, 335, 337⁷. **b**, Residues in which exposed side chains were altered by site directed mutagenesis in the homologous mouse IgG2b antibody. Open circles mark mutants which are lytic. The carbohydrate attachment residue, Asn 297 is shaded: Glu 318, Lys 320 and Lys 322 are marked with solid circles. The following residues were altered to alanine: Ser 239, Lys 248, Ile 253, Asp 265, Ser 267, Asp 270, Gln 274, Glu 283, His 285, Gln 290, Glu 294, Asn 297, Lys 317, Glu 318, Lys 320, Lys 322, Lys 326, Glu 333, Thr 335, Ser 337 and Lys 340. Pro 331 was mutated to glycine, and Glu 235 to leucine, but residue 235 is not shown since it is located in a disordered region of the structure²³.

Methods. Sheep red blood cells were washed three times in PBS and resuspended at 10^9 cells ml⁻¹. NIP-kephalin²⁵ (5 μ l, 100 mg ml⁻¹) was added to 1 ml washed cells and incubated at 37 °C for 2 h. The cells were washed three times in complement fixation diluent (CFD, Oxoid UK) and resuspended to 10^8 cells ml⁻¹. Serial dilutions of antibody were made in CFD in flat bottomed microtitre wells in 50 μ l. Guinea pig serum (50 μ l, 1/10 dilution) and derivatized cells (100 μ l) were added. The plate was incubated at 37 °C for 30 min and the wells scanned at 690 nm using a Titertek Multiscan. The results were converted to % lysis using a standard curve for red cell lysis, and plotted against antibody dilution. In the wild-type antibody and most of the mutants above, 3 μ g ml⁻¹ antibody was sufficient for 50% lysis. However, the mutants Asn 297 $\rightarrow$ Ala, Glu 318 $\rightarrow$ Ala, Lys 320 $\rightarrow$ Ala and Lys 322 $\rightarrow$ Ala were inactive.

(Thr 318 and Arg 320) and rat IgG2a (Thr 318) (ref 1; for rat isotypes, M. Bruggemann personal communication). To evaluate the effect of these substitutions in mouse IgG2b, we made three point mutants Glu 318 $\rightarrow$ Thr, Lys 320 $\rightarrow$ Arg and Lys 322 $\rightarrow$ Arg. Their properties in complement lysis or binding to C1q proved identical to wild type (Table 1). To further dissect the requirements of the C1q site, three more mutants Glu 318 $\rightarrow$ Val, Lys 320 $\rightarrow$ Gln and Lys 322 $\rightarrow$ Gln were made. These proved non-lytic. In the direct C1q binding assay, the mutants Glu 318 $\rightarrow$ Val and Lys 322 $\rightarrow$ Gln bound poorly to C1q. However the Lys 320 $\rightarrow$ Gln mutant had an affinity for C1q only slightly weaker than wild type (Table 1), indicating that good C1q binding may be insufficient for lysis, and that perhaps a precise orientation of C1q is required (see below). Thus for effective lysis, arginine



or lysine (but not glutamine or alanine) are required at residues 320 and 322, and glutamic acid or threonine (but not valine or alanine) at residue 318. This suggests an electrostatic interaction with C1q at residues 320 and 322 and perhaps a hydrogen bond at residue 318.

All antibody isotypes sequenced to date possess the C1q binding motif, or a closely related motif which is effective in binding C1q when it is transplanted into the mouse IgG2b antibody. Clearly there must be further determinants for lysis. For example, antibody isotypes with short hinges and low segmental flexibility are non-lytic¹⁵, suggesting that the interaction of C1q with the motif is sterically blocked due to close approach of the Fc by the Fab arms¹⁶, or that the interaction of C1q and antibody requires an exact alignment for lysis and therefore requires some flexibility *per se*. Since the motif is conserved even in the non-lytic isotypes, these isotypes may have a role, perhaps in the modulation of lysis. For example the presence of the human IgG4 isotype in immune complexes with the human IgG1 isotype appears to inhibit C1q binding and lysis¹⁷.

The residues 318–322 are located in a short linear section of polypeptide chain, and a peptide mimic of this region should inhibit lysis by competing with antibody for binding of C1q. Indeed simple organic and inorganic ions¹ and peptides^{18,20} inhibit the binding of antibody to C1q or C1 (C1q, C1r, C1s complex). We designed a peptide based on the motif above, with alanine residues as spacers and blocked N- and C-termini: AcAla-Glu-Ala-Lys-Ala-Lys-Ala-NH₂. This peptide inhibited lysis at 170 μ M, at a concentration similar to free antibody (Fig. 3). We have not formally identified the stage in the complement cascade at which the peptide acts, but peptide inhibitors with a similar constellation of side chains, that is, ... Asn-Ala-Lys-Thr-Lys ..., were found to act before C2 conversion in an Augener C1 consumption assay^{18,20}.

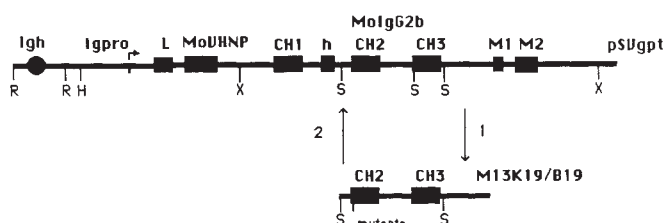


Fig. 2 Construction of mutants of the mouse IgG2b antibody. The vector for expression of the heavy chain¹⁰ was partially digested with *Sac*I to release a fragment containing the CH₂ and CH₃ exons of the mouse γ 2b gene. This was subcloned in the vector M13K19 (step 1), and the internal *Sac*I site removed using a synthetic mismatch primer. Each of the mutants was made on this template using site directed mutagenesis with synthetic primers of between 14 and 21 nucleotides in length, with central mismatches of between one and three nucleotides²⁶. The *Sac*I fragment was excised and recloned in the expression vector (step 2). Igh, IgG heavy chain enhancer; Igpro, IgG heavy chain promoter; exon L, heavy chain signal peptide; exon MoVHNP, heavy chain variable domain; exons CH1–3, heavy chain constant domains; h = hinge; M1 and M2, membrane domain. Restriction sites: R, *Eco*RI; H, *Hind*III; S, *Sac*I; X, *Xba*I. After transfection of the myeloma J558L, secreted antibody was purified on a 1 ml column of NIP-Sepharose²⁷.

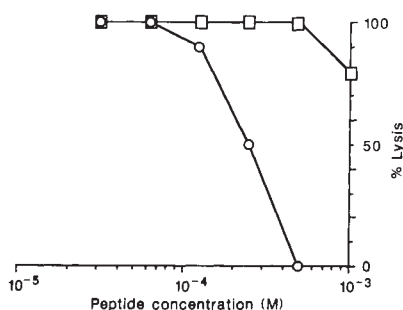


Fig. 3 Inhibition of complement lysis by synthetic peptides. Circles mark lysis for the peptide AcAla-Glu-Ala-Lys-Ala-Lys-Ala-NH₂ (based on the motif), and squares mark a control peptide Ser-Asp-Glu-Lys-Ala-Ser-Pro-Asp-Lys-Tyr, which includes both acidic and basic residues.

Methods. Lysis of sheep red blood cells was as in Fig. 1. A checkerboard assay (ref 28: mouse IgG2b antibody titrated in the first dimension and guinea pig complement source in the second) was used to determine the lowest antibody concentration giving complete lysis after 30 min incubation. Peptide was purified by reverse phase HPLC and the composition checked by amino acid analysis. Freeze dried peptide was weighed and dissolved in CFD to a concentration of 4 mM and serial dilutions of the peptide made into 200 μ l NIP-derivatized red cells (5×10^7 cells per ml), guinea pig serum (1/40 dilution) and antibody (7 μ g ml⁻¹), and incubated and scanned as in Fig. 1. An apparent I_{50} value of 170 μ M compares with an I_{50} value of 75 μ M (or 150 μ M per CH₂ domain) reported for monomeric IgG¹⁸.

In conclusion we have identified a C1q binding motif, Glu318, Lys320, Lys322 in the CH₂ domain of the antibody, and a simple peptide based on this sequence is able to inhibit complement lysis. The charged nature of the C1q binding motif, the requirement for basic residues at 320 and 322, the ionic strength dependence of C1q binding, and the peptide inhibition data all suggest an electrostatic interaction. There are charged regions in each of the three chains in C1q heads, for example the sequence Glu 198-X-Asp 200-Y-Lys 202 in the A chain²¹ which might interact with the motif as an antiparallel β strand. The C1q binding motif is a common sequence found in many proteins, and we suggest that recognition of this common motif in an array (for example in viruses and bacteria) directly by the

multiple heads of C1q is an ancient defence mechanism. For example, retroviruses can be lysed directly by C1q in the absence of antibody²². Later in evolution, the antibody may have appeared as an adaptor, avoiding the requirement for the motif to be written directly on the pathogen.

We thank D. R. Burton and M. S. Neuberger for discussions, M. Bruggemann and H. Waldmann for the sequences of rat isotypes in advance of publication and R. Valerio for the peptides.

Received 8 February; accepted 15 March, 1988.

- Burton, D. R. *Molec. Immun.* **22**, 161-206 (1985).
- Reid, K. B. M. & Porter, R. R. *Biochem. J.* **155**, 19-23 (1976).
- Hughes-Jones, N. C. & Gardner, B. *Molec. Immun.* **16**, 697-701 (1979).
- Metzger, H. in *Contemporary topics in Molecular Immunology* **7**, 119 (Plenum, New York, 1978).
- Yasmeen, D., Ellerson, J. R., Dorrington, K. J. & Painter R. H. *J. Immun.* **116**, 518-522 (1976).
- Colomb, M. & Porter, R. R. *Biochem. J.* **145**, 177-183 (1975).
- Burton, D. R. *et al. Nature* **288**, 338-344 (1980).
- Alexander, R. J. & Steiner, L. A. *Molec. Immun.* **17**, 1263-1273 (1980).
- Bedouelle, H. & Winter, G. *Nature* **320**, 371-373 (1986).
- Neuberger, M. S. & Williams, G. T. *Phil. Trans. R. Soc. A* **317**, 425-432 (1986).
- Koide, N., Nose, M. & Muramatsu, T. *Biochem. biophys. Res. Commun.* **75**, 838-844 (1977).
- Nose, M. & Wigzell, P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6632-6636 (1973).
- Leatherbarrow, R. J. & Dwek, R. A. *FEBS Lett.* **164**, 227-230 (1983).
- Leatherbarrow, R. J. *et al. Molec. Immun.* **22**, 407-415 (1985).
- Oi, V. T. *et al. Nature* **307**, 136-140 (1984).
- Isenman, D. E., Dorrington, K. J. & Painter, P. H. *J. Immun.* **114**, 1726-1729 (1975).
- van der Zee, J. S., van Swieten, P. & Aalberse, R. C. *Clin. exp. Immun.* **64**, 415-422 (1986).
- Prystowsky, M. B., Kehoe, M. & Erickson, B. W. *Biochemistry* **20**, 6349-6356 (1981).
- Boackle, R. J., Johnson, B. J. & Caughman, G. B. *Nature* **282**, 742-743 (1979).
- Lukas, T. J., Munos, H. & Erickson, B. W. *J. Immun.* **127**, 2555-2560 (1981).
- Reid, K. B. M., Gagnon, J. & Framton, J. *Biochem. J.* **203**, 559-569 (1982).
- Cooper, N. R., Jensen, F. C., Welsh, R. M. & Oldstone, M. B. A. *J. exp. Med.* **144**, 970-984 (1976).
- Deisenhofer, J. *Biochemistry* **20**, 2361-2370 (1981).
- Brunhouse, R. & Cebra, J. J. *Molec. Immun.* **16**, 907-917 (1979).
- Weltzien, H. U., Matthiessen, H. P., Meyer-Delius, M., Zimmermann, F. & Rude, E. *Molec. Immun.* **21**, 801-810 (1984).
- Carter, P., Bedouelle, H. & Winter, G. *Nucleic Acids Res.* **13**, 4431-4443 (1985).
- Neuberger, M. S., Williams, G. T. & Fox, R. O. *Nature* **312**, 604-608 (1984).
- Campbell, A. M. *Monoclonal Antibody Technology* (Elsevier, Amsterdam, New York, Oxford, 1984).
- Leatherbarrow, R. J. & Dwek, R. A. *Molec. Immun.* **21**, 321-327 (1984).

CORRIGENDUM

Multiple potassium-channel components are produced by alternative splicing at the *shaker* locus in *Drosophila*

Thomas L. Schwarz, Bruce L. Tempel,
Diane M. Papazian, Yuh Nung Jan & Lily Y. Jan
Nature **331**, 137-142

IN Fig. 3b in this letter, a base was missed in the sequence of ShD1. An additional G should be present as the 48th nucleotide from the 5' end. Thus the correct sequence beginning at nucleotide 44 should read: GAGCGGCTGC. By shifting the reading frame of the extreme 5' end, a stop codon is introduced at nucleotides 42-44. Thus one or more of the five AUG codons that are present in this splicing variant and were identified as potential translational start sites are likely to serve this function and the predicted ShD protein is likely to be of approximately the same size as the other predicted products of the locus.

In addition, in Fig. 1 the locations of the mutations Sh^{K82a} and T(1; 3)Sh^{LC} were inadvertently shifted 5 kb to the left of the positions to which we had mapped them¹. They still fall within the same intron as shown. A second graphics error has shifted the location of intron 11 in Fig. 3a one nucleotide to the left of its correct position.

Finally, the authors regret having omitted a reference to Goldman *et al.*² who demonstrated that a gene for a subunit of a putative neuronal nicotinic acetylcholine receptor is alternatively spliced. Thus alternative splicing may be a mechanism by which diversity is produced for many receptors or channels in both vertebrates and invertebrates.

The authors thank Dr Olaf Pongs for pointing out the sequence error.

- Papazian, D. M., Schwarz, T. L., Tempel, B. L., Jan, Y. N. & Jan, L. Y. *Science* **237**, 749-753 (1987).
- Goldman, D. *et al. Cell* **48**, 965-973 (1987).

Table 1 Affinity of human C1q for mutant mouse IgG2b antibodies

Mutant	Lysis	Affinity for C1q (nM)
Asn 297 $\rightarrow$ Ala	-	31
Glu 318 $\rightarrow$ Ala	-	>300
Glu 318 $\rightarrow$ Thr	+	11
Glu 318 $\rightarrow$ Val	-	110
Lys 320 $\rightarrow$ Ala	-	>300
Lys 320 $\rightarrow$ Arg	+	11
Lys 320 $\rightarrow$ Gln	-	13
Lys 322 $\rightarrow$ Ala	-	>300
Lys 322 $\rightarrow$ Arg	+	11
Lys 322 $\rightarrow$ Gln	-	>300
wild type	+	11

The affinities of the mutants Glu 235 $\rightarrow$ Leu, Ile 253 $\rightarrow$ Ala, Gln 290 $\rightarrow$ Ala, Pro 331 $\rightarrow$ Gly, Glu 333 $\rightarrow$ Ala, Thr 335 $\rightarrow$ Ala and Ser 337 $\rightarrow$ Ala for C1q were also determined and lie in the range 9-12 nM. A 10% suspension of Affigel 701 (Bio-Rad) polyacrylamide beads were derivatized with 4-hydroxy-3-iodo-5-nitrophenacetyl- ϵ -amino-*n*-caproic acid succinimide ester (Cambridge Research Biochemicals, UK) in 0.125 M sodium carbonate, at 37 $^{\circ}$ C in the dark for 24 h. Excess reagents were removed by successive centrifugation and decantation, and the NIP-cap-Affigel support stored as a 7% suspension in PBS containing 0.02% sodium azide at 4 $^{\circ}$ C in the dark. The support had a capacity for mouse IgG2b of about 10 mg ml⁻¹. The C1q binding assay used ¹²⁵I-labelled human C1q and was based on ref. 29.

Science and the New Journalism

Walter Gratzer

"I WENT to the R: Society, where were Tryals with Sir E Newtons Burning-glasse: which did strange things as to mealting whatever was held to it in a moment: one of the most difficult was common Slate, which lasted longer than Iron, Gold, brasse, Silver, flint, brick etc which it immediately mealted, calcined and Vitri-fied: The Glasse was composed of 7 round burning glasses of about a foote diameter, so placed in a frame as to cause all their Sun-beams to meate in one focus onely". Thus John Evelyn in 1705, setting a respectable standard for science writers to come.

Consider now this recent passage from a great English daily. It is the Science Correspondent himself who speaks: "Experiments in transplanting genes, made both on the laboratory bench and the natural ones removed from other organisms, are running into difficulties because they are being rejected by the host on which they are imposed. . . . Since the manufacture of some new pharmaceutical or other chemical is dependent on getting the maximum number of organisms to sympathize with their respective compound, the rejection process is very serious". Clear? Good, because there is plenty more in the same vein. Who was it who said that to be intelligible is to be found out? (John Evelyn would also not have informed us, as an American newspaper did recently, that "the knee is the Achilles heel of the leg".)

Wet sand

Would reading an account of a football match, a Royal outing or a natural disaster equally make one feel as though one were being beaten about the head with a sock full of wet sand? Why did the editor permit such an abomination against the canons of his craft to remain on the page? The answer surely must be that science barely qualifies as news. As an attention-grabbing headline "W-Boson Discovered" may rank with "Small Earthquake in Peru". There is of course the counter-argument, however unfashionable, that one does not necessarily have to concern oneself with the preferences of the average reader (or writer) of English newspapers. Dylan Thomas expressed the principle like this:

*Few understand the works of Cummings,
And few James Joyce's mental slummings,
And few young Auden's coded chatter;
But then it is the few that matter.*

Attempts have, to be sure, been made to package science for the lower end of the market. Prototypic examples include the works of Paul de Kruif, of which *Microbe*

Hunters is the best known. De Kruif had been a microbiologist — or, as he always called himself, a microbe hunter — at the Rockefeller Institute and supplied the technical background for Sinclair Lewis's great scientific novel, *Martin Arrowsmith*. De Kruif adhered to the *Gott-im-Himmel* breathed - Robert - Koch - staring - at - the - sectioned - *Leberwurst* - on - the - microscope - slide - But - these - can - only - be - the - fatal - anthrax - bacilli school of historical writing which is now all but unreadable. Yet *Microbe Hunters* appears to have inspired any number of young people of an earlier generation to take the vows and dedicate themselves to a life in the laboratory.

De Kruif's style of reportage has enjoyed a resurgence in recent years. Here, for instance, is a sample, taken at random from a book by a successful science journalist, published by an American university press: "Chromosomes, chromosomes, you could always find Elmer Lustwinkel at the table, sorting through hundreds of black-and-white enlargements of chromosomes. Long straight chromosomes; short stubby chromosomes; chromosomes that looked like bow ties, black ants, and licorice twists; and all wearing stripes known as banding patterns. Slowly, deliberately, *ploddingly*, Elmer Lustwinkel would cast his big soft eyes over each photograph studying the shapes and patterns". (I have changed only the name; the italics are the author's.) The flow of bilious prose exerts a curiously unsettling effect as the familiar terrain takes on an unearthly hue.

Scientists themselves are now of course beginning to generate their own myths. The great journalist and aphorist, Karl Kraus, opined that wars were started and history was made by politicians telling lies to journalists and then believing what they read in the newspapers. This is also the style of some of the less attractive members of the scientific profession — those with press agents, for instance — and the relation between science and journalism has in consequence become an uneasy one: the journalist seeks to exploit the eager hyperbole of the scientist and the scientist the untutored credulity of the journalist. But a new breed of science journalist is emerging, more committed and better informed, often indeed with a professional background in science.

The New Journalism is a term apparently invented by Tom Wolfe. As he defines it, it has two main components. The first is the quality of the writing; what Wolfe calls the aesthetic dimension can elevate it to a literary form, which, he holds, is now thriving at the expense of the novel. The

second element is that the journalist writes as an insider, a participant, however transiently, in the events that he describes. He rides, like Hunter Thompson, with the Hell's Angels, he runs onto the field, like George Plimpton, with the Detroit Lions, or, like Tom Wolfe, he heroically nibbles canapés of Roquefort, rolled in crushed walnuts, with the Black Panthers at Leonard Bernstein's fund-raising party. In the end of course the trust and hospitality on which this *modus operandi* depends is generally betrayed, and for this a heavy price is sometimes exacted: Hunter Thompson was left on the floor of a roadhouse in California, spitting teeth and blood and pondering the *mot juste* with which to conclude his best-seller.

Candour

The New Journalism has at last advanced into science. Jim Watson's memoir, *The Double Helix*, can be seen as a precursor; it was undeniably a participant's view and its candour gave wide offence. There are now some admirable practitioners of the craft (practically all, it has to be admitted, in America). They include Nicholas Wade (*The Nobel Duel*), William Broad (*Star Warriors*) and, more recently, Stephen Hall (*Invisible Frontiers*) and Gary Taubes (*Nobel Dreams*). All know how to write compelling prose and all are formidably well briefed and pertinacious. The boundary between journalism of this calibre and serious contemporary history has become blurred. Stephen Hall's account of the expensive scramble by competing research groups to achieve the first expression of a recombinant human gene, and Taubes's chronicle of Carlo Rubbia's (and CERN's) pursuit of the mystical Nobel, illuminate the darker side of modern science, hold up the mirror to Narcissus, record stirring chapters of history in all their triumph and squalor and, I cannot doubt, ultimately do a major service to science and the community.

Not the least of the qualities that such writers must possess are endurance and fortitude. A man who has spent a year with a tape-recorder in the mephitic caverns of CERN, or in the Californian salt-mines where clones are cloned, will carry marks to compare with those of Hunter Thompson on the roadhouse floor, or George Plimpton, plastered into the mud of the stadium in Detroit. One is not after all put into this world for pleasure alone. □

Walter Gratzer is in the MRC Cell Biophysics Unit, Kings College London, 26–29 Drury Lane, London WC2B 5RH, UK.

Hawking's timely story

Don N. Page

A Brief History of Time: From the Big Bang to Black Holes. By Stephen W. Hawking. Bantam, New York: 1988. Pp. 198. \$18.95 (United States), \$24.95 (Canada). To be published in Britain on 16 June.

STEPHEN Hawking is a remarkable scientist. Despite being afflicted with amyotrophic lateral sclerosis, he has become the greatest gravitational theorist since Albert Einstein. For example, in work with Roger Penrose, Hawking showed that Einstein's classical theory of gravity, general relativity, implies that our Universe had a beginning. He went on to show that black holes normally can only grow in area, though in 1973 he made the surprising discovery that quantum-mechanical effects can allow black holes to emit radiation and thereby shrink. Since that time, Hawking has focused on applying quantum mechanics to gravity and to the Universe as a whole.

Now Hawking has taken time aside from his research on the forefront of cosmology and quantum gravity to write a popular book on the subject. The result is a brief but highly readable account of the development of scientific ideas about the Universe from Aristotle's to his own. Without using any equations beyond $E = mc^2$, Hawking gives a lucid account of special and general relativity, quantum mechanics, elementary particle physics, black holes and their emission, the Big Bang, inflation, the state of the Universe, the arrow of time and the unification of physics. Most of the first three-quarters of the book covers ideas that are fairly well accepted, but the final quarter concentrates upon more speculative proposals.

Because there has been a veritable big-bang explosion of popular books on physics and cosmology recently, one may question the need for another one. This can be answered in two ways. First, it is probably fair to say that *A Brief History of Time* is the first popular book on cosmology written by a scientist of such stature as Hawking since Steven Weinberg's *The First Three Minutes* appeared eleven years ago.* Indeed, the two books are complementary, because Weinberg wrote from the perspective of an elementary particle theorist, with an emphasis on non-gravitational forces and the contents of the Universe, whereas Hawking writes from the perspective of a general relativist, with an emphasis on gravity and the geometry of the Universe as a whole. Both books are intended for people without a scientific education, but Hawking's will probably be slightly easier to read,

*Steven Weinberg's book has just been reissued by Basic Books — see the review on the facing page.

because it is somewhat less detailed than Weinberg's and relies more on qualitative rather than quantitative reasoning.

The second need met by *A Brief History of Time*, in its final quarter, is through the discussion on a popular level of the 'no-boundary' proposal for the quantum state of the Universe, which Hawking made in qualitative form at a conference in the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Stephen Hawking — boundless ideas.

Vatican in 1981 and in mathematical form in a paper with James Hartle in 1983. (The idea, roughly, is that one does a Feynman path integral over complex spacetimes without boundary.) This proposal, if correct, would specify which solution of the dynamical equations of a 'unified field theory' or a 'theory of everything' describes our particular Universe. The precise properties of the proposed solution involve infinitely difficult computations which, of course, can never be carried out, but approximate calculations using highly simplified models for the dynamical equations already suggest that the no-boundary proposal may explain the

mysteries of why our Universe is so large, is so nearly homogeneous and isotropic on the largest observed scales, has considerable structure on smaller scales, and shows such a strong thermodynamic arrow of time. It is highly doubtful that these observed properties could ever be explained by a purely dynamical theory of everything, because the equations that we know so far (admittedly not the final ones) allow solutions for the Universe that do not have these properties.

The no-boundary proposal is scientifically controversial for several reasons. First, as Hawking emphasizes, it is just a proposal and cannot be deduced from other principles; indeed, there are now some competing proposals that are not mentioned in the book. Second, for technical reasons related to the attractiveness of gravity, the proposal has not really been made precise for the dynamical equations for which it has been formulated, quantum general relativity coupled to matter fields. (This problem of the 'indefiniteness of the gravitational action' was presumably considered too complicated to be mentioned.) Third, we do not yet know the form of the ultimate dynamical theory of everything to which the proposal should be applied. Hawking argues that we already know several features it should have, but if the ultimate theory does not even fundamentally involve spacetime as a basic entity (a distinct possibility in superstring theory, for example), the no-boundary proposal would have to be reformulated. Finally, although the preliminary results are encouraging, there have not yet been enough detailed predictions calculated from the proposal to test it adequately against observations.

Although many of these caveats are noted in the book, there is the danger that Hawking's enthusiasm for his proposal may lead the less-cautious reader to become more convinced of its correctness than there is yet evidence to warrant. Nevertheless, because of the significance of the proposal if it is correct, its exposition in this timely book is valuable for enabling the public to ponder some of its implications.

A Brief History of Time provides Hawking with an opportunity to present at first hand some of his own philosophical and theological conclusions from the no-boundary proposal. After discussing what he considers to be the most common view of God's activity — that God started off the Universe and then let it evolve without intervention — Hawking objects that if the Universe actually has no boundary, and hence no beginning, "What place, then, for a creator?". However, this objection does not apply to the Judaeo-Christian view, that God creates and sustains the entire Universe rather than just the beginning. Whether or not the Universe has a beginning has no rele-

vance to the question of its creation, just as whether an artist's line has a beginning and an end, or instead forms a circle with no end, has no relevance to the question of its being drawn.

Of course, if the no-boundary proposal is correct, it does have interesting implications for the beauty, elegance and simplicity of the Universe God did create. Hawking draws the conclusion that God "had no freedom at all to choose initial conditions", but this is debatable. When I was defending Hawking's proposal to a small group of gravitational theorists in 1982, Bryce DeWitt expressed this view by saying, "You do not want to give God any freedom at all". However, Karel Kuchař quickly rejoined, "But that's His choice". In other words, even if we correctly hypothesize which state God chose

for the Universe, that would in no way eliminate the freedom He may have had in making that choice. Choosing the no-boundary state and then actually carrying out the immense task of the creation of the Universe in this state is a far cry from Carl Sagan's claim in his introduction to the book of "nothing for a Creator to do".

So although the final quarter of *A Brief History of Time* is most interesting, it is also controversial. Nonetheless, the book should be acquired by anyone seeking to learn about some of the latest ideas and speculation in cosmology, and to gain familiarity with the insights that have sprung from the extraordinary mind of Stephen Hawking. □

Don N. Page is Professor of Physics at The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

... and Weinberg revisited

David Layzer

The First Three Minutes: A Modern View of the Origin of the Universe. By Steven Weinberg. Basic Books: 1988. Pp.198. Hbk \$16.95; pbk \$7.95. In Britain the book is out of print.

Big Bang, Big Bounce: How Particles and Fields Drive Cosmic Evolution. By Iosif L. Rozental. Translated by Juri Estrin. Springer-Verlag: 1988. Pp.129. Pbk DM58, £22, \$35.

AFTER eleven years, *The First Three Minutes* remains, in my opinion, the best introduction to cosmology for non-mathematical readers. Galileo, the first and most successful popularizer of cosmological theories, pinpointed the central difficulty of such an enterprise when he remarked (in another context) that the book of nature is written in the language of mathematics. Like poetry, mathematics always loses something in translation. Yet Weinberg manages not only to capture some of his subject's deepest ideas in clear, lively prose but also to communicate a personal vision of what science is and how it grows. This new edition is a reprint of the original text and of an afterword added in 1982.

Two main themes are intertwined in the book: the history of the early Universe and the history of modern cosmology. Weinberg pays special attention to the historical antecedents of Arno Penzias and Robert Wilson's accidental discovery, in 1965, of the cosmic radiation background, and he devotes an entire chapter to the question of why astronomers didn't look for (and find) the background radiation earlier.

The text of *The First Three Minutes* and

the seven-page afterword of 1982 are themselves interesting historical documents. In 1976, the 'standard model' of the early Universe, which interprets the radiation background as the fossil remnant of a primordial fireball, seemed firmly entrenched. It explained the black-body character of the background radiation and it made predictions of the primordial abundances of helium and deuterium that seemed to be consistent with astronomical evidence. It did not, however, explain why the Universe is clumpy or why it contains roughly a thousand million photons for every nucleon. Clumpiness and the photon/nucleon ratio had to be put in 'by hand' — postulated as initial conditions. But most cosmologists were confident that these problems would eventually be solved. In 1976 Weinberg saw only one potential threat to the standard model, "one great uncertainty that hangs [over it] like a dark cloud": although the radiation background is now uniformly bright in all directions, there is no direct evidence that it was uniform and isotropic at very early times, as the standard model assumes.

By 1982, when Weinberg wrote his afterword, the picture had changed. New astronomical data and refined analyses of old data implied that if the standard model is correct, only a small fraction (now estimated at one per cent or less) of the mass in the Universe resides in ordinary matter. The bulk of the mass must reside in particles for whose existence there is as yet no independent experimental evidence. Meanwhile, Weinberg and other particle physicists, following up an early suggestion by Andrei Sakharov, had constructed a new theoretical foundation for the standard model. They conjectured that the Universe initially contained equal numbers of particles and antiparticles, and argued that the slight excess of particles needed to account for the observed photon/baryon ratio could have been pro-

duced by baryon-nonconserving processes during a critical early period of the cosmic expansion. Baryon-nonconserving processes are permitted by unified theories of the strong and electroweak interactions. These theories also allow neutrinos to have small but finite mass. Could the missing mass be in the form of massive neutrinos? In 1982 the answer seemed to be yes.

Since 1982 the picture has changed again. Massive neutrinos have been ruled out as carriers of the missing mass, and no adequate replacement for them has emerged. The simplest unified theory of the strong and electroweak interactions has been found to conflict with experimental lower limits on the lifetime of the proton. Alan Guth's inflationary hypothesis, put forward in 1981 and described briefly in Weinberg's afterword, has solved some the problems — notably the magnetic-monopole problem — but it has created others. Increasingly stringent observational limits on small-scale angular variations in the brightness of the radiation background have made it increasingly difficult to reconcile the hypothesis of a primordial fireball with the observed clumpiness of the astronomical universe. And the standard model has split, at its base, into a dozen or so competing hypotheses, each more speculative than the last.

One of these hypotheses — that the Universe is made up of non-interacting 'metagalaxies', each with its own physical constants and laws — forms the central theme of *Big Bang, Big Bounce*, a book intended to introduce readers who "have taken courses in physics and mathematics at the undergraduate or graduate level" to some elementary concepts of particle physics and cosmology. In his preface, Rozental praises Weinberg's book, and his text is intended as a kind of sequel to *The First Three Minutes*.

I found the sections on particle physics clear and interesting. The sections that deal with observational and theoretical cosmology are less satisfactory. For the most part they are superficial and perfunctory, and some of them — for example, the discussion of 'Newtonian cosmology', a topic that Weinberg treats very well — are seriously misleading. As for the translation, it preserves the flavour of the original Russian so successfully that one is hardly aware that one is reading English. □

David Layzer is the Donald H. Menzel Professor of Astrophysics at Harvard University, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

• Another classic example of popular exposition in cosmology, Arthur Eddington's *The Expanding Universe*, which was first published in 1933, has recently been re-issued in paperback by Cambridge University Press. Price is £7.95, \$12.95.

Inside a physicist

John Stachel

H.A. Kramers: Between Tradition and Revolution. By M. Dresden. *Springer-Verlag*: 1987. Pp. 563. DM 120, \$74.95, £45.

THE Dutch theoretical physicist Hendrik Antonie Kramers is probably known to more physicists as an initial than as a name. How many realize what lies hidden behind the modest 'K' of the WKBJ approximation method in quantum mechanics? A smaller number will be familiar with one or another of such phrases as Kramers theorem or the Kramers–Kronig relations, but his name is hardly a household word among the current generation of physicists. Yet, as

major factors leading Kramers to turn away from active participation in work of the quartet just when it was about to culminate in the definitive formulation of the new mechanics.

Kramers left Copenhagen in 1926 to accept the chair of physics in Utrecht, becoming the dean of the Dutch physics community after the deaths of Lorentz and Ehrenfest. His life and scientific activity after the move are treated sketchily in Part Three, except for a chapter of over 100 pages devoted to his work on quantum electrodynamics. Like Pauli, Kramers was very critical of Dirac's approach. He felt that the answer to the problems of the quantum theory lay in a careful re-examination of the classical theory of the interaction of electrons and radiation. Dresden shows that his faithfulness to the classical legacy of Lorentz's electron

of its author in several ways. Dresden is a distinguished physicist from Holland, who studied with Kramers before going to the United States in 1939, and who met him again several times after the war. He is clearly very involved, both intellectually and emotionally, with the subject and his milieu (he spent 11 years working on this book), and is well acquainted with many others who figure in the story. The book is based on a close reading of Kramers's writings, both technical and popular, and of a host of other sources in physics and the history of physics. As might be expected from a physicist of Dresden's eminence, the explanations of Kramers's work are masterly and will be indispensable for those interested in the early history of quantum mechanics and of quantum electrodynamics. But Dresden attempts much more than exposition. He

AIP Niels Bohr Library

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Kramers and contemporaries — the famous line-up at the Fifth Solvay Conference in 1927. Kramers is fourth from the left in the second row.

Max Dresden sets out to demonstrate, he has a claim to be remembered along with that distinguished trio consisting of his friends, Wolfgang Pauli and Werner Heisenberg, and his mentor, Niels Bohr, who contributed so decisively in the early 1920s to shaping the emerging quantum mechanics, and through it the course of physics over the past 60 years.

The depiction of the interrelationships between Kramers and the other members of this quartet, and a masterly discussion of some of its consequences for Kramers's research, form the central core of the book. After some introductory scientific background in the first four chapters, the remainder of Part 1 gives a quick introduction to the quartet and a summary of their accomplishments during the decade (1916–1926) that Kramers was in Copenhagen as Bohr's assistant. The second part is a detailed survey of Kramers's work during this period, which culminate in two rewarding but technically demanding chapters on the Bohr–Kramers–Slater (BKS) theory and the Kramers–Heisenberg dispersion theory. The author establishes the importance of Kramers in the events that led Heisenberg to formulate matrix mechanics. He suggests that Kramers's depression after the collapse of the BKS theory, in which he had placed such exaggerated hopes, as well as fear of competition with Heisenberg, to whom Kramers felt intellectually inferior, were

theory was a source of both strengths and weaknesses in Kramers's work. After the Second World War, Kramers's approach, with its emphasis on the need for mass renormalization, started to receive widespread recognition among physicists searching for a way to understand the newly discovered Lamb shift. But Kramers again turned away from active participation in the further development of the theory that led to the now classic Feynman–Schwinger–Dyson relativistic version of quantum electrodynamics.

In the final part of the book, Dresden attempts to assess Kramers's personality and scientific style, his social and religious commitments (or lack of them), and his self-image. A lot of additional biographical material is presented here, so that a reader trying to form a connected picture of Kramers's life has to do a good deal of skipping around. The author explains that he was forced to cut down a manuscript of some 1,400 printed pages, which may account for certain imbalances and gaps in the treatment of various periods in Kramers's life, as well as the scattering of information on some topics throughout the book. (The circumstances of Kramers's marriage, for example, are discussed at several widely separated places.) The index is very weak and does not make it any easier to gather together such information.

This biography bears the distinct stamp

tries to explain the intra-psychoic reasons for the succession of 'near misses' in Kramers's career — cases in which he just failed to make some major discovery. He also speculates on why Kramers never worked on certain problems that he was eminently qualified to tackle.

These portions of the book lean very heavily on a number of personal interviews with members of Kramers's family, close friends and fellow scientists, who would presumably not have spoken so freely to someone in whom they did not have full confidence. Kramers's children made available his diaries, date books and private poems. This material enables Dresden to provide a number of remarkable insights into his subject's inner life.

The picture that he presents is a gloomy one. Dresden sees an essential continuity between certain basic traits manifested in Kramers's closest interpersonal relations, in his relations with his fellow scientists and even in his relationship to science itself. An inability to commit himself completely to anyone, friend or wife, or to any cause, religious or political, showed itself early in his life and deepened as he grew older. He feared any major decision, because it foreclosed all alternative paths. His strong attraction to and talent for such varied fields as music, literature and science made the choice of a career particularly difficult for the young man. His first mentor, Paul Ehrenfest, felt that this

trait would prevent him from succeeding in physics. Kramers also had a horror of interpersonal conflicts, and would go to extraordinary lengths to avoid them. He regarded this as a family characteristic, telling a young relative "A Kramers has to fight to fight". He was well aware that his inability to fight for his convictions hindered him scientifically.

Perhaps the most remarkable example is Kramers's unpublished anticipation of the Compton effect. The idea occurred to him in 1921, but he allowed Bohr, who at the time was a vigorous opponent of the photon concept, to talk him out of publishing his work. Indeed, Bohr was so effective in converting Kramers to his point of view that Kramers became an even more intemperate opponent of the photon than Bohr himself. Compton later received a Nobel prize for his work, a prize which it is reasonable to suppose Kramers would have shared had he predicted the effect two years before the experiment. It is sad to learn that Kramers's life, like that of many other outstanding scientists, was embittered by his failure ever to receive a coveted Nobel. One is left with the impression of a brilliant, upright and conscientious man, torn by incessant internal conflicts, constantly tormented by feelings of personal and professional lack of fulfilment, who maintained external control over his feelings at a great cost to himself.

How accurate is this picture? It is hard for an outsider to know. Dresden freely admits that the book gives "my understanding of Kramers' life". The wealth of intimate detail about Kramers that Dresden provides relies so heavily on personal interviews (Dresden himself notes the "soft" character of this information) that it is difficult for others to assess the evidence until the interviews (which I hope were taped), as well as Kramers's personal papers, are made available to others. Information that would seem important to an assessment of Kramers's psychological development is not always provided. (His father turns up remarried when Kramers is in his mid-twenties, for example, without any indication of what happened to his mother.)

We must be immensely grateful to Dresden for having broken with the genteel tradition of scientific biography by writing a fascinating book that leads us into the forbidden territory of Kramers's personal and psychic life. However, one sometimes feels the need for a more expert guide through this territory, the sort of superlative guide that Dresden proves to be through the world of Kramers's physics. □

John Stachel, 745 Commonwealth Avenue, Boston, Massachusetts 02215, USA, is Professor of Physics at Boston University and Editor of the Collected Papers of Albert Einstein, the first volume of which appeared in 1987.

Jump in quantum history

Nicholas Kemmer

The Historical Development of Quantum Theory. Vol. 5 Erwin Schrödinger and the Rise of Wave Mechanics. Part 1 Schrödinger in Vienna and Zürich 1887–1925; Part 2 The Creation of Wave Mechanics: Early Response and Applications 1925–1926. By Jagdish Mehra and Helmut Rechenberg. *Springer-Verlag: 1987. Pp. 980. Part 1 DM98, £36, \$54. Part 2 DM148, £55, \$79.95.*

QUANTUM theory was born with this century and needed a quarter of it to reach maturity as quantum mechanics. That theory soon gained almost universal acceptance as the key to a much-deepened understanding of inanimate nature.

As Jagdish Mehra told us at the start of this multi-volume work, he began preparing it just about another quarter-century later. For him, the time was then most auspicious; nearly all those who, from 1925 onwards had taken part in creating and developing quantum mechanics, were still active, not to mention some of *their* teachers. Mehra established happy contacts with, it would seem, nearly all these people. A large part of the work that Mehra then began is now in print. Since 1970 he has had Helmut Rechenberg as partner.

Mehra displays a special ability to convert copious records of formal interviews and informal exchanges into lively accounts that ring very true. These are not just incidental but rather quite central to the structure of the first four volumes of the work, which are not strictly chronological in order. They split into parts presenting the ideas that relate roughly to the work of Planck, Bohr, Pauli, Heisenberg and Dirac. Embedded into the very readable technical sections there are biographies and pen pictures of these men and descriptions of their backgrounds. Naturally the work of many other people is described, some of whom receive similar, though shorter treatment. Also, as an introduction, figures from the prehistory of quantum theory appear in the setting of their achievements, with very detailed documentation for each one.

The reader-interest of different sections of these volumes inevitably varies. For me, the later technical chapters deal with topics and publications I remember from long ago, while earlier parts invite study for the twists and turns of how it was before my time. As for the vast amount of background information and biographical detail, one must be grateful that it exists.

Now, after a five-year gap, a further two-part volume of this work has been

published. It is dedicated to Erwin Schrödinger, and appears at a time when the centenary of his birth has just been celebrated. Schrödinger was undoubtedly a great physicist and a remarkable person. Why, however, was he not covered together with the 'greats' of the previous volumes? That would have been appropriate chronologically — and indeed his name is not absent from those earlier pages. Schrödinger receives special treatment because, within the history of quantum theory, he holds a quite unusual position. He was a most distinguished scientist who worked on many aspects of theoretical physics, but until 1926 his contributions to quantum theory were quite modest. In that year he published a series of papers that immediately transformed ideas and methods in quantum theory quite profoundly. In due course this work gained him a Nobel prize. However, because the interpretation of his 'wave mechanics' by the great majority of his colleagues differed drastically from his own, he soon parted company with them. Only his wave equation flourished within quantum mechanics.

This volume is clearly meant both to continue the series and to tell Schrödinger's story. Given the authors' general intent to provide maximum background material, it could hardly be the former without the latter. There was thus a great amount of further background material to collect beyond Mehra's earlier work (which had included interviews with Schrödinger), explaining the time gap between the appearance of the previous volumes and this one. Schrödinger left a huge amount of archival material behind which the authors not only use but also list in detail, and in addition much has been assembled on his Viennese background and some on Zürich, where his irruption into quantum theory took place.

Description of Vienna in the late nineteenth century begins the work. We are told some history of the multinational Habsburg Empire and are introduced to its many-sided culture, including a number of its great figures. Schrödinger enters as an inheritor of great traditions — a polymath growing up in a highly cultured family. Every event, place or person mentioned is the subject of detailed documentation.

The reader learns of many men who influenced Schrödinger in his youth. Perhaps the most relevant of them is Ludwig Boltzmann, because statistical mechanics became one of Schrödinger's main interests. (Boltzmann is first introduced alongside Ernst Mach, and we hear incidentally of their great debate on the acceptability of atoms as real.) In 1912 Schrödinger began to be known to theoretical physicists through his prolific publications on many topics. Space forbids discussing them here, as Mehra and Rechenberg do

New volumes in the SCIENTIFIC AMERICAN LIBRARY

Eye, Brain, and Vision

David H. Hubel

In *Eye, Brain, and Vision*, David Hubel — a pioneer in the physiology of the brain and co-recipient of the 1981 Nobel Prize for Physiology and Medicine — takes us on a fascinating tour at the frontiers of neuroscience, where he and other researchers are conducting breakthrough experiments to probe the complex processes of vision.

January 1988 240 pages
150 illustrations Board £14.95

Molecules

Peter Atkins

'This is undoubtedly the most beautiful chemistry book ever written. The idea was to bring together superb photographs and high-quality molecular graphics, and explain the link between them. The idea is not new... but never has it been done so well, and for the non-chemist. Peter Atkins manages it in *Molecules* in a way that not only informs but stimulates and delights... This is descriptive chemistry triumphant!' *New Scientist*

1987 256 pages
290 illustrations Board £14.95

The Science of Structures and Materials

J. E. Gordon

This volume examines the strength of structures and materials from a scientifically eclectic viewpoint. Renowned for his contributions to materials science, J.E. Gordon shows us how the same mechanical principles that underlie the strength of natural structures like tendons and muscles apply to man-made structures like skyscrapers, ships, and aircraft.

May 1988 219 pages
189 illustrations Board £14.95

Sand

Raymond Siever

Sand presents a geological view of a common material, showing how one can use its properties to learn a great deal about the geological processes operating at the Earth's surface and the physics and chemistry underlying them. Readers will enhance their understanding and appreciation of beaches, dunes, river bars, and in general, the workings of the Earth.

July 1988 240 pages
190 illustrations Board £14.95

The book publishing arm of *Scientific American*

W H Freeman

20 Beaumont Street,
Oxford OX1 2NQ

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Back in Vienna — Schrödinger in 1956.

in interesting detail. Most of these papers were concerned with properties of matter in bulk, but with digressions, for example into problems of colour vision.

In the present context, Schrödinger's interest in general relativity is noteworthy. It led to his study of the ideas of Weyl, which linked the theory with electromagnetism. There, a line integral (of the four-potential) featured centrally. In 1922 Schrödinger remarks in a note that this integral, taken round a hydrogen Bohr orbit, is proportional to its quantum number! Apart from this prophetic shot, single atoms had been hardly mentioned up to then. Even as his great year of 1926 was approaching, it was not the atom but problems in statistical mechanics that occupied Schrödinger. We read of an interesting discussion between Planck and Einstein on the statistics of the ideal gas, in which they sought to agree on how to count states in an assembly. Schrödinger joined in and was encouraged by Einstein to write a paper. Einstein also told him of Bose's ideas, and then of de Broglie's recent startling suggestion that, for counting purposes, the gas molecules should be represented by waves (de Broglie's work was described in an earlier volume). Evidently it was this replacement of particles by waves in gas statistics — a counterpart to Planck's introduction of the quantum into the statistics of radiation 25 years earlier — that stimulated Schrödinger to study de Broglie's idea. It seems that the part of the latter's paper on the hydrogen spectrum, which relates closely to Schrödinger's 1922 paper, was not where wave mechanics started! In fact Schrödinger seems to have forgotten that three-year-old work of his own. At this point the great idea was not yet fully formed. Debye is reported to have asked Schrödinger

Popperfoto

whether he had considered the *wave equation* for the new 'matter waves'. When he did so, he first had to examine and discard the relativistic form of the equation. Then came success. The paper, giving the right result for the hydrogen spectrum, appeared on 27 January 1926.

When, shortly afterwards, Schrödinger beat Pauli with his proof that the new 'Göttingen' quantum mechanics and his 'wave mechanics' were *mathematically* the same thing, the great leap forward was complete. Here the first part of the authors' story ends. It is also roughly the end of the first half-volume. It might have been entitled *The Early History of Erwin Schrödinger*.

The second half could be called *The Early History of the Wave Equation*. At this point, abruptly, Schrödinger's equation transformed the character and the rate of progress of research in quantum theory. It divides roughly into two streams. On the one hand, much needed doing before the full scope of the wave equation could be understood. The quantum mechanics initiated by Heisenberg had relied for its general formulation on the use of the classical theory of Hamilton. For the understanding of Schrödinger's equation a similar bridge to a 'classical limit' was found in the Hamilton-Jacobi equation. On the other hand, many concrete problems could be solved by wave mechanical methods and are reported in detail by Mehra and Rechenberg. Thus we read of Heisenberg's explanation of the helium spectrum and then of the concept and applications of Fermi statistics. Schrödinger contributed to both kinds of studies. There are careful and detailed accounts of all this work. However, no single piece of work here reported pinpoints the paradox in the Schrödinger story more than Born's treatment of scattering, in which he made the decisive pronouncement that the wave function does not directly describe any observable physical entity, but is a means of defining *probabilities* of the outcome of observations. This interpretation was forever unacceptable to Schrödinger, an opinion which was shared by Einstein but only a few others.

There are, apparently, more volumes of *The Historical Development of Quantum Theory* to follow. Whether they will record the prolonged debates for and against the accepted probabilistic interpretation of quantum mechanics, we do not know. In any case, we are told by the pioneers of today to look beyond the four-dimensional world of particles for some 'reality' in many dimensions, with objects such as strings inhabiting it. Was Schrödinger so wrong not to believe in the survival of Born's picture? □

Nicholas Kemmer, 35 Salisbury Road, Edinburgh EH16 5AA, UK, is Emeritus Professor of Mathematical Physics in the University of Edinburgh.

Looking up at the Lick

David H. DeVorkin

James Lick's Monument: The Saga of Captain Richard Floyd and the Building of the Lick Observatory. By Helen Wright. Cambridge University Press: 1987. Pp. 231. £25, \$32.50.

Eye on the Sky: Lick Observatory's First Century. By Donald E. Osterbrock, John R. Gustafson and W. J. Shiloh Unruh. University of California Press: 1988. Pp. 295. \$25.

BUILT in a young and vigorous California, at a time when visual refracting telescopes dominated astronomical technology, the Lick Observatory at its opening in 1888 was potentially the most powerful astronomical tool in existence. Its patron was one of the most unusual characters of post-gold rush California; its backers and builders were not astronomers, though they consulted astronomers world-wide in their planning; its instrument and dome makers, save for the firm of Alvan Clark and Sons, had never ventured anything of even remotely comparable scale; and its first director was a far from stable fellow. It is remarkable that the enterprise turned out as well as it did.

The founding years of the observatory are recalled with romantic beauty and admirable accuracy in Helen Wright's *James Lick's Monument*. By focusing on Captain Richard Floyd, a son of the Confederacy, and on his sense of divine service to science, Wright provides a well-documented and valuable glimpse into an important episode of patronage, politics and pride in late-nineteenth-century American science, including the display of national chauvinism that made it impossible for Howard Grubb, the famed Dublin instrument-maker, to bid successfully to build the telescope or dome.

The early years of life at Lick were very tough. Even during its construction, the director-designate, Edward Singleton Holden, was at odds with Floyd, the builder of the observatory and president of the Lick Trust. And once Holden gained control of the magnificent new telescope, his bright, highly capable staff did not find his direction to their liking.

The pioneer astronomical spectroscopist of Tulse Hill, Sir William Huggins, once quoted the words of a fellow astronomer upon hearing of the staff problems at Lick Observatory: "What do you expect when a few men are shut up together on the top of a mountain, and only one big telescope?" (William Huggins to George Ellery Hale, 10 December, 1892, Yerkes Observatory Library). Although that

quotation is not to be found in *Eye on the Sky*, Donald Osterbrock, John Gustafson and W. J. Shiloh Unruh do deftly exploit the wealth of archival material held in the Mary Lea Shane Archives of the Lick Observatory to tell the story not only of the building of the observatory, but of its first painful years under Holden and the happier decades since. Indeed, this rich archive, collected together and established by Mary Shane, is the heart of both works reviewed here.

Though Osterbrock *et al.* treat Holden with more sympathy than does Wright, both paint similar portraits of the dynamics of the first years of observatory life. It took only a short time for Lick's 36-inch refractor, the largest of its type in the world and situated in the clear and dry air of 4,300-foot Mount Hamilton, to shift the geographical centre of astronomical observation to the West. For example, Huggins, locked in mortal combat with Norman Lockyer over the chemical composition of nebulae, regarded observations from Lick as the arbiter in their battles, though both ultimately interpreted the data as they saw fit.

Holden was a bitter disappointment. His best staff — James Keeler, Sherbourne Wesley Burnham and Edward Emerson Barnard — all left within six years. Keeler returned as director only after Holden was ousted, and he managed in the few years before his untimely death in 1900 to establish observational astro-

physics at Lick.

Keeler's legacy was carried on in a programmatic fashion by William Wallace Campbell. Campbell embarked on a long-term stellar radial velocity programme, and encouraged his staff to maintain other cataloguing efforts, such as Robert Grant Aitken's double star measures. Programmatic observing was then the *raison d'être* of astronomy.

At a late point in their history, Osterbrock, Gustafson and Unruh discuss briefly the transition from general programmatic observing to individual staff-initiated projects at Lick in the 1950s. As an example of the limitations of the programmatic approach, they provide a litany of solar eclipse spectroscopic expeditions mounted by the observatory which failed to anticipate future needs. Wouldn't it have been nice, the authors muse, if Donald Menzel had been on the staff earlier than 1926, soon enough to encourage Campbell to provide sensitometry marks on his plates! Menzel required line intensities in addition to wavelengths. A student of astrophysics under Henry Norris Russell of Princeton, Menzel may well have appreciated the need for calibration earlier, but his whole programme was a result of the very recent work of Meghnad Saha on ionization equilibrium. A 'Donald Menzel' was impossible *circa* 1900.

Thus the authors give the impression that the Lick Observatory was particularly

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

conservative in its programmatic tradition, making the transition late and only with some pain. This may well be so, but Lick was representative of an age and far from unique in the often difficult and bitter process of changing to an emphasis on individual projects informed by astrophysical theory.

To Osterbrock and his co-authors, Lick Observatory was vital only when it exploited properly its high mountain site and clear stable air. Thus they hail Keeler's upgrade of the 36-inch Crossley reflector and his creative use of it for observational astrophysics. The Crossley remained Lick's primary tool for direct photography and nebular spectroscopy, but it was soon overshadowed by George Ellery Hale's 60-inch reflector on Mount Wilson. As other observatories built larger reflectors, Lick's status continued to slip; only with the completion of the 120-inch Shane reflector, and the recruitment of staff able to explore both the observational and conceptual edges of modern astrophysics, did it fully revive.

For the past 20 years, Lick Observatory has been the Department of Astronomy at

the University of California at Santa Cruz. Osterbrock and company provide a valuable narrative chronology of events that shows how and why this happened. Although at times the story is laboured — more a 'who did what and when' than an intellectual journey — the staff and programmes of the observatory and how and why they changed over time are candidly and vividly portrayed.

My only complaint about *Eye on the Sky* is that it is devoid of direct citations to sources. Writing a popular work, the authors did not wish to "clutter the text with references" (p. 274), feeling that sources cited in the aggregate in bibliographical notes should be sufficient for those motivated to dig further. More serious readers will find the practice frustrating. But that apart, both books do credit to the centenary of the Lick Observatory and are well worth reading for the complementary perspectives they provide. □

David H. DeVorkin is Curator of Astronomy at the National Air and Space Museum, Smithsonian Institution, Washington, DC 20560, USA.

Here, there and everywhere

Desmond King-Hele

Infinite in all Directions. By Freeman J. Dyson. *Bessie/Harper & Row: 1988. Pp.321. \$19.95.*

FREEMAN Dyson is among the few scientists who are comfortable roaming infinitely in all directions, and his new book lives up to its exacting title. He takes us on an exhilarating roller-coaster ride that runs from the farthest reaches of boundless space into the nutshell-world of the subatomic, and from life's origins to its present travails and faraway future. The book springs from the Gifford Lectures he gave at Aberdeen, Scotland, in 1985 under the title "In Praise of Diversity", but the material has been extended and rewritten.

Dyson begins by skipping round the physical Universe, telling us of events and ideas that appeal to him, and then offers three subtle and detailed chapters on the origin of life and its possible future. After these essays on the infinite possibilities of the Universe, he turns to our "unbounded responsibilities" as custodians of life on a small planet, where we have "a heritage to cherish" and "a destiny to earn". Looking back at his own roots, he sees himself as an intellectual descendant of Richard Hakluyt, the sixteenth-century promoter of exploration; of William Blake, the prophet of freedom; and of Jules Verne,

who set him "on the road to America and from there onward to Uranus".

Darting here and there as the title promises, Dyson next focuses on the idea that "quick is beautiful" in technology today: being able to react to change within three to five years is the key to success. A new and growing technology, he believes, "is usually small, quick and agile"; an overripe technology "cannot regain its lost youth". The virtues of the quick and the small re-emerge in the chapter entitled "Science and Space", and are contrasted with the vices of massive projects where the design has to be fixed ten years before the spacecraft operates. Dyson compares NASA's Hubble Telescope unfavourably with ESA's Hipparcos spacecraft. "Committees do harm merely by existing", he says, because the big and costly projects are always over-emphasized. Also committees work by extrapolation of current trends; but this is not effective for more than about ten years ahead, because "neither cleverness nor stupidity is predictable" and either may alter the climate of opinion irrevocably. So engineers need to dream, and the chapter "Engineers' Dreams" offers the committees some nightmares. Dyson's own dream is for a space mission to Uranus in 2016 by an Astrochicken, a symbiosis of plant, animal and electronic components. Launched as an egg, it will mature at its planetary destination and manoeuvre itself with the aid of a Bombardier beetle's chemical jet. Imagine a committee deliberating on that proposal!

From biological spacecraft, Dyson flips back to ponder power politics on Earth.

He likes the example of Austria, non-nuclear and neutral, and advocates the Austrianization of Germany. He hopes for a non-nuclear world, but not world government; and he is even-handed in his discussions of Star Wars and nuclear winter. These varied subjects are linked by his beliefs (apparently serious and not tongue-in-cheek) "that the universe is constructed according to a principle of maximum diversity" and that the laws of nature "are such as to make the universe as interesting as possible".

Whatever we may think of the maximum-diversity principle as a world-view, there is no denying its importance throughout this scintillating book that gives the lie to the literary wallahs who complain that scientists cannot write. And only a scientist could run over the Universe in a manner so easy and relaxed. Dyson is to be congratulated on making science interesting for an audience wider than scientists, and on writing much better than most of the literary pundits.

Some of the sparkle arises because Dyson cannot resist climbing out on a limb, and no one will agree with all that he says. For example, he candidly confesses to believing in a God "who grows with the universe and learns as it develops". For those who need a Deity, this concept of God as a super-Freeman-Dyson may perhaps be more attractive than many of the other Gods on offer in the world today. Dyson also goes out on a limb in trying to argue that science and religion are similar: this seems a libel on science, for religion has been and still is at the root of a large fraction of the world's wars, whereas science has not directly caused any, and only a few can be attributed to technological arrogance.

Differences of opinion on these wider issues are inevitable, and Dyson should not be blamed for his frank speaking. A more serious complaint, I think, is that his enthusiasm and personal touch — which together animate the book — do sometimes lead him to swallow theories rather uncritically, especially if their originators have visited him at Princeton. To mention one example, comets may possibly come from an unobservable "Oort Cloud", and this hypothetical cloud may have a "loss-cone" to explain periodicities. There is also a remote possibility that the Sun has a companion star (they call it Nemesis). But even the first two of these hypotheses are not "firmly based", as he says they are: they may soon evaporate like a puff of phlogiston. Enthusiasm brooks no ifs and buts. This book rides on its enthusiasms and, though the ride is pleasant, we often need to add our own ifs and buts. □

Desmond King-Hele is a Deputy Chief Scientific Officer at the Royal Aircraft Establishment, Farnborough, Hampshire GU14 6TD, UK. Among other books he is author of Satellites and Scientific Research (1960) and Erasmus Darwin and the Romantic Poets (1986).

A cosmic swindle

Joseph Silk

The Symbiotic Universe: Life and Mind in the Cosmos. By George Greenstein. *William Morrow:1988. Pp.255. \$18.95.*

The Cosmic Blueprint. By Paul Davies. *William Heinemann, London/Simon & Schuster, New York:1987-1988. Pp.224. £12.95, \$17.95.*

The Quickening Universe: Cosmic Evolution and Human Destiny. By Eugene T. Mallove. *St Martin's Press:1988. Pp.268. \$18.95.*

ONCE upon a time there was a flea, who lived on the back of a dog called Rover. The flea thrived, and grew as big as a pea, for Rover's fur was the perfect environment, neither too hot nor too cold, too dry or too moist, too dark or too light. Then one day, the dog's mistress purchased a flea collar for her pet. From that day on, the world was a very harsh place for the flea, who no longer had a moment's peace. Sadly, the flea was forced to jump, and landed by a lamp-post, where he soon froze to death.

The moral of this story is that although a canine universe may be eminently well suited for the existence of fleas, fleas are not inevitable. Some dogs have them, some do not. It is equally presumptuous to believe that man's existence is inevitable, as advocated by proponents of what has become known as the strong anthropic principle. The Universe does provide a suitable environment for man to exist, but the flea's precarious existence should caution us that even this weak anthropic argument is more tautology than teleology. Rover's fur provides a breeding ground for fleas quite independently of whether or not fleas exist. The Universe is isotropic and homogeneous and long-lived and matter-dominated, all ingredients more or less necessary for man's existence, because of events a fraction of a second after the Big Bang. Cosmologists continue to debate why those events happened, and precisely how they proceeded, but their conclusions are not going to account for the origin of life. Nor should the procession of unlikely events that heralded the evolution of life be allowed to get cosmologists off the hook: they still have much explaining to do.

George Greenstein, in his new book *The Symbiotic Universe*, lends a sympathetic ear to this sort of teleological tattle. The complexities of nature offer a glimpse of a grand designer, or rather, of a grand design, with the designer absent, essentially in hiding. Existence of a watch implies that there was a watchmaker: we are urged to take the logical next step, that existence of a flower implies that a cosmic flowermaker must once have plied his

craft. From the flower, it is simple to inquire about the origin of a sentient being, and presto, we have proven the existence of a deity.

There is, of course, a flaw in this logical progression, and I wish that all budding grand designers would take due note. We know how to make a watch, hence we correctly infer the existence of a watchmaker. Our cleverest botanists do not know how to create a daisy, and we are still farther from constructing a chimpanzee. Hence we can conclude nothing about their designers: we lack the information necessary to make any inference. A reductionist approach fails utterly to account for a solitary daisy. The laws of physics, applied to the individual molecules, could succeed only with negligible probability. Some additional ingredient is needed. Is this necessarily where the theologians enter?

In *The Cosmic Blueprint*, Paul Davies takes the rational approach. Unpredictability is a vital element of many non-linear systems. Push the simple pendulum just so far, and it no longer exhibits its predictable swing. The turbulent patterns, whether of cream stirred in coffee or of cumulus cloud formations, are infinitely more complex than their basic elements (coffee, cream or water vapour). The theory of the behaviour of simple dynamical systems provides us with an understanding of how non-linearity can trigger seemingly chaotic patterns. It is when one goes from physical to biological systems that an intriguing question arises. Is biological complexity mechanistic in origin, or do we need to go beyond the laws of physics and chemistry to account for the diversity of a living system? Perhaps the non-linearities that arise when a complex

broth of primordial amino acids is suitably processed will induce sufficiently non-random mutations and syntheses to lead to otherwise highly improbable primitive life forms.

The real challenge is yet to come. Is a sentient being no more nor less than a highly complex non-linear dynamical system? Certainly, the implausibility arguments of Crick, Hoyle *et al.* which purport to demonstrate the need for an extraterrestrial origin of life are no longer viable. They must be abandoned in the face of chaotic behaviour and the ensuing unpredictability of dynamical systems. Whether consciousness, which is an extreme form of organization, can arise in this manner cannot be answered, if we do not know the capacity of matter for self-organization.

The Quickening Universe by Eugene Mallove treads much of the ground covered by Greenstein and Davies. It adds more, including chapters on the search for extraterrestrial life and on the origins of music. The result unfortunately reads like the collection of articles that spawned the book, being mostly a mish-mash of anecdotes and scientific morsels, with little depth or coherence. Mallove is a freelance writer, but his prose style does not compare with the lyrical writing of Greenstein, an astrophysicist.

There is suddenly a surfeit of scientific popularizations on aspects of cosmology, and *The Quickening Universe* is one book that can safely be set aside. Instead, for a good read, take *The Symbiotic Universe*; for insights into creation, try *The Cosmic Blueprint*. □

Joseph Silk is a Professor in the Department of Astronomy, University of California, Berkeley, California 94720, USA.



Dangerous work — radium-dial painters unaware of the health risks they faced. The picture is taken from Health Risks of Radon, produced by the 4th Committee on the Biological Effects of Ionizing Radiations (BEIR IV). The committee was established in response to a request from the Environmental Protection Agency and the Nuclear Regulatory Commission to assess the risks associated with internally-deposited alpha-emitters. Published by National Academy Press, Washington, DC, the book costs hbk \$42.50, pbk \$32.50.

Back to reality?

V. Paul Marston

Observation and Objectivity. By Harold I. Brown. Oxford University Press: 1987. Pp. 255. £22.50, \$29.95.

Objective Knowledge. Edited by Paul Helm. Inter-Varsity Press, Norton Street, Nottingham NG7 3HR, UK: 1987. Pp. 192. £10.95.

THEOCHARIS and Psimopoulos's attack on philosophers of science, which appeared in *Nature* of 15 October last year, certainly stimulated general interest in questions of objectivity and scientific truth — but did it rightly reflect recent thinking in these areas?

The logical empiricism which dominated the 1940s–1950s emphasized logic, language and a search for neutrality of observation. Actual science was used to illustrate a supposedly logically constructed system defining objective truth (and denying any other kind). Karl Popper to some extent shared this 'logical' approach, but hinted at the radical criticism that was to come in books by, for example, Hanson (1958), Polyani (1958) and Kuhn (1962). These authors began by looking at actual historical science rather than logical constructions of it, and developed ideas (seen earlier in Whewell and Duhem amongst others) that all observation is 'theory-laden'.

By the late 1960s logical empiricism was dead. But what was to replace it? If all observation is theory-laden, then surely whether science has 'progressed' depends on your point of view, and scientific theories are useful instruments rather than 'truths'? Kuhn himself recoiled from the idea, but Feyerabend developed it and some extremists concluded that science is simply an ideology, with no special claim to truth.

Other philosophers of science, however, such as Hacking, Putnam, Shapere and Suppe, whilst likewise beginning from actual science and accepting that theory is implicit in observation, have sought (in a variety of ways) to develop ideas which preserve and redefine objectivity, scientific realism and scientific 'progress'. Though such realist movements have their critics (for example Laudan and Van Fraassen) they have largely set the agenda of debate since the late 1970s. This makes ironical the plea in Theocharis and Psimopoulos's article that work should be done "putting forth adequate definitions of such fundamental concepts as objectivity, truth, rationality and the scientific method", as it is exactly what such 1980s figures have been attempting.

Professor Brown reflects the general trend. His book *Perception, Theory and Commitment*, published in 1977, con-

tained a long epitaph for logical empiricism followed by a Kuhnian analysis of perception with paradigms and normal and revolutionary science. *Observation and Objectivity* hardly mentions Kuhn, and its index has no entry for 'paradigm'. It advocates the scientific realism of the school of Dudley Shapere, considering the function of scientific theory, the nature of knowledge and scientific observation, and the meaning of objectivity.

Brown's treatment shares some of the weaknesses and incompleteness of Shapere's ideas. Thus the twin ideas of 'domain' and 'theory' (introduced by Shapere in 1969 as "fundamental conceptual tools for illuminating the nature of science") remain obscure. Popper, Kuhn and Lakatos all told us how to recognize a scientific discipline, and why (for example) astrology wasn't one. Shapere's 'domains' remain vague, and Brown's frequent reference to them clarifies little. To say, for example, that "the notion of a domain encapsulates all items that researchers deal with in a particular field" (p.23) and "different norms are appropriate in different domains" (p.32) offers little help in recognizing them.

Brown states that one of his main aims is to argue that the study of how we get knowledge is itself a domain: "epistemological claims have the same status as theoretical claims in science" (p.13). The correct way to obtain knowledge is learned by studying how scientists in fact did it in the past: "epistemology is a second order science" (p.33). The problem here is that in assuming that particular sciences *have* got knowledge, we surely need to know what knowledge is? Why should all our *present* concepts not eventually be classed as mistakes — as 'ether' is now? If so, any present epistemology would have been based on a study of mistakes rather than on genuine instances of knowledge acquisition.

Similar circularity of argument exists in Brown's drawing of parallels between the old idea that perception is 'caused' by something 'out there' and a belief in the reality of scientific non-observables. Surely a question about the 'reality' of any *particular* physical entities is meaningful only if we have a *prior* commitment to the general existence of a world 'out there'? Brown's refutation of 'instrumentalism' is little more than an assertion that it is not in fact what scientists do, and he gives us no reason for confidence that all our present concepts will not eventually be reduced to mere 'instruments for prediction' just as Newtonian mechanics has been (p.15).

If, however, Brown's realism is not yet convincing on such points, this simply indicates need for further discussion. His book does successfully sum up in a readable manner some of his school's 'realist' thinking on the nature of observation and science, and contains some useful case

analyses. I hope that not only the philosophically minded, but many thoughtful scientists whose gut reaction may be 'realist', will want to ponder upon the questions and problems modern scientific thinking poses on these issues.

Objective Knowledge is a compendium of articles from academic practitioners of science, social science, linguistics and divinity, each of whom addresses the question of objectivity in his or her discipline. The contributors take an evangelical Christian standpoint — committed (as mainstream evangelicalism has always been) both to the inspiration of the Bible and to the general reliability of mainstream science. The high religious commitment of so many scientists who built our present world-view should (if for no other reason) make us wish at least to consider this viewpoint seriously.

In a book published recently, one might have wished for more reference to the new realist movements, of which the authors say almost nothing. They do, however, come to grips with the relativism of 'Kuhnism', Polyani (a longstanding fascination of Walter Thorson's) and parallel figures such as Mannheim. They analyse how 'theory-ladenness' relates to the Christian view of a 'real' created world.

The book contains a refreshing breadth of vision in our world of increasing specialization by tackling 'objectivity' across very different disciplines. It well represents the current general thinking of Anglo-American Christians who take this particular position (such as those of the American Scientific Affiliation), and as such is a useful library addition. □

V. Paul Marston is a Senior Lecturer in the Faculty of Science, Lancashire Polytechnic, Corporation Street, Preston PR1 2TQ, UK.

Biology and the Bible

Two recent books on science and theology are *God and Evolution* by R. J. Berry (Hodder, £6.95) and *Biblical Creation and the Theory of Evolution* by D. C. Spanner (Paternoster, £6.95). The authors champion the classic Christian viewpoint that has been increasingly under attack over the past decade by supposed Biblical 'literalists', who argue for a 'young Earth' created in 144 hours.

Berry and Spanner are biologists who see nature and the Bible as 'God-given' and in ultimate harmony, but base their science on observation not on theology. The treatments overlap, but Berry focuses more on the history of the interaction between science and theology, contemporary evidences and 'creationism', while Spanner concentrates on the Biblical meaning of creation and its compatibility with modern science. Both books usefully illustrate how a belief in a Biblical world-view can be consistent with modern empirical biology. V. Paul Marston

Why are 'they' not like 'us'?

Stephen Jay Gould

Racial Theories. By Michael Banton. Cambridge University Press:1987. Pp.181. Hbk £22.50, \$44.50; pbk £6.95, \$12.95.

DEATH and taxes, the old quip proclaims, are life's only inevitabilities. W.H. Auden retained death, dropped taxes and added love in listing his pair of the only subjects worth a writer's attention. To this select list of human ineluctables, I propose that we add race. We are unavoidably fascinated by variation, and cannot help wondering how or whether our behaviours and abilities relate to physical differences among us.

This fascination with differences has persisted through all the vicissitudes and revolutions of Western culture. Othello and Jesse Jackson both face the same problem of acceptability, though all terms and interpretations have altered during the centuries separating the Doge's Venice and Reagan's America. Banton's book treats these changing contexts as a taxonomy of theories about race from the scientific revolution of Newton's age to our own.

Racial Theories contains five chapters, one for each category of Banton's taxonomy: race as lineage, type, subspecies, status and class. All theories attempt to answer the same question: why are 'they' not like 'us', but the terms of resolution could not be more different, or more deeply embedded in changing social contexts and changing scientific knowledge. The lineage theory argued: "because they belong to a different genealogical group that acquired disparate characters". The type theory retorted: "because they were made differently *ab initio*". The subspecies theory proclaimed: "because they have differentiated by evolution from a common point of origin". Status and class are more modern entries from the social sciences into a domain that had been debated largely in biological terms.

The fascination of this book and subject lies in the complexities and studied ambiguities of Banton's taxonomy. How do these five theories relate; how can they be usefully subdivided and contrasted? Taxonomies are mirrors of thought, theories about the organization of knowledge. They are not just pigeon holes or book-keepers' devices for ordering the objective facts of nature. If we can develop a coherent taxonomy of attitudes towards race, we will take the most important of all steps towards understanding.

Banton's taxonomy has no single and obvious structure. It may be viewed and organized in many ways — as intended, I think, by Banton who wants to show us that taxonomies are human constructions,

not inevitable logics, or facts of nature. We might view the sequence of five as a simple series of steps in growth of sophistication of knowledge — from pre-evolutionary views of biology (lineage and type) to Darwin's biological world (subspecies) to a modern recognition that social forces must be given their due, if not their dominance (status and class).

But this cannot be entirely right because Banton's sequence of five is not sequential in time. His five theories include two synchronic pairs — contrasting views that

suggests, for example, that status vs class embodies a large set of disparate divisions in disciplinary style (micro vs macro), philosophical predisposition (Kant vs Hegel) and personal politics (liberalism of the comfortable academic variety vs radicalism).

I enjoyed this basic feature of the book's organization. It taught me something about the uses of taxonomic ordering. It made me think about the subtle interactions of our ordering minds and nature's patterns. It provided an active metacommentary upon an important aspect of scientific history.

But *Racial Theories* also contains a debilitating flaw. It is so full of errors (in the biological sections, where I claim competence to judge) that one simply loses confidence. Too many great ideas pass by too quickly, often in single paragraphs based entirely on secondary

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Racial collage — "If we can develop a coherent taxonomy of attitudes towards race, we will take the most important of all steps towards understanding".

actively (and contentiously) competed. Lineage and type do not replace each other. They are Banton's terms for the classical debate between monogeny (descent of all races from a single Adam and Eve, with differentiation according to lineage), and polygeny (creation of races as separate species, or types) — the greatest argument of pre-Darwinian anthropology. Similarly, status and class represent a division now actively debated in the social sciences.

Perhaps, then, we should view the five theories as synchronic, with a primary division between three biological and two social views. But this won't work either, for some sense of progress surely lies embedded in Banton's sequence. The creationist concept of type is truly superseded by the evolutionary view of subspecies, while status and class dilute the crude biological determinism of nineteenth-century racism. Moreover, when synchronism does prevail, the basis for distinction is multifaceted. Banton

sources, themselves constructed from cardboard accounts in the heroic mode of steady progress in objective knowledge. I do realize that most of these errors, in themselves, are trivial, but you must excuse the parochiality of an American who blanches upon being told that Thomas Jefferson was his country's second president (he was third) — just as Englishmen would rightly laugh with derision if I wrote that Henry VI had eight wives, or that Edward VIII was Victoria's rotund son.

Other errors are central and more serious, showing that Banton does not grasp many of the biological concepts he discusses. He writes, for example (p. 83), that sexual behaviour in the Malagasy mouse lemur shows "that at the very root of primate evolution there was an extremely loose kind of relationship between the sexes". But a modern species at the 'bottom' of a conventional primate list is not an ancestor — and the lemurs, as a large group of varied species, display

an enormous range of sexual behaviour, with no single version ranking as indubitably ancestral for all primates. Moreover, for a man so interested in taxonomy, Banton might learn the basic Linnaean rules. He writes, for example, that calling the herring gull by a trinomial, *Larus argentatus argentatus*, denotes a decision not to designate subspecies. "The species name is repeated to show that it is not divided at the subspecific level" (p. 4). Exactly the opposite is true. Repetition of the trivial name expresses a decision to divide the species into subspecies: the repeated trivial name only indicates that this particular subspecies has been designated as the 'type' or namebearer for the species.

Social context and advancing knowledge have fuelled all the transitions and backings and forthings among Banton's theories. What will happen in the future as genetic evidence indicates an astonishing similarity among 'races' (a product of the probable origin of *Homo sapiens* as a branching event in Africa less than 250,000 years ago, a geological yesterday) — while social divisions of first and third world continue to deepen, thus weakening the very bonds that biological knowledge is now illuminating. □

Stephen Jay Gould is a Professor in the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

THE MANAGEMENT OF AIDS PATIENTS

Edited by DAVID MILLER, JONATHAN WEBER and JOHN GREEN

The Management of AIDS Patients is the first comprehensive guide to the practical clinical management of patients with AIDS or HTLV III infections. The book avoids the sensational aspects of the disease, offering solid advice and information for all people involved in patient care.

The editors and many of the contributors come from St Mary's Hospital, London one of the foremost centres in Britain for the treatment of AIDS patients. The knowledge and experience of these experts has been combined to provide a much-needed book for doctors, nurses, dentists and indeed for health-care professionals everywhere.

1986

200pp	Hardback	0 333 40465 3	£30.00
	Paperback	0 333 40466 1	£12.95

Please order this title from your bookseller — or in case of difficulty from Felicity Davie, Macmillan Press, Houndmills, Basingstoke, Hants RG21 2XS.

M MACMILLAN PRESS

Jacobs' bequest

T.C. Whitmore

The Tropical Rain Forest: A First Encounter. By Marius Jacobs. Translated and edited by Remke Kruk *et al.* Springer-Verlag: 1988. Pp. 295. Pbk DM64, \$39.95, £24.

Complete fugues could be composed from the leaves of *Shorea*, varying in size, thickness, venation and hairs; with counter-melodies of fruits and stipules; a single fermenta for a particularly large flower; and a tremolo for an aberrant calyx. And from time to time the recurrent melody of the flowers' exquisite scent.

Shorea is a big genus of giant rain-forest timber trees. As a series of word-pictures which conjure up the majesty of tropical rain forests and the awe they inspire, Jacobs's book, from which the above quotation is taken, has few modern equals. It is written in the great tradition of nineteenth-century accounts of exploration and tropical natural history, quoting some of their more purple passages and adding some memorable new ones. It is further enhanced by the numerous magnificent illustrations, some original, many borrowed. The book was first published in 1981, in Dutch. The author died in 1983 while still translating and enlarging it. Colleagues have completed the task and added some new material.

Tropical rain forests, as everyone now knows, have been and are being hit hard. Modern machinery, rising populations, insatiable overseas markets for cheap (because undervalued) timber, the chance for politicians and businesses to become rich, all contribute. Jacobs was appalled by what he saw happening, and his book is a sustained, finely written polemic against destruction for short-term gain in the name of development. It contains some well-aimed barbs against the lack of vision of all who condone or engage in such 'development', and is by far the most eloquent and fully argued such book. The chapters on how biological knowledge has developed and is built up, and on the information we have on the New World rain forests, make excellent introductions to their subjects. That on Africa is less adequate (for example White's vegetation map and commentary are scarcely mentioned). The emphasis throughout is on the Eastern tropics with which Jacobs was best acquainted.

As its subtitle tells, the book is a 'first encounter', and as such truly conveys the flavour of the forests. But it is not successful as an ecological text. Part of the problem is that only 20 of the 400 references cited are more recent than 1981, so several newer developments are left out. Our understanding of forest mineral nutrition



From *The Tropical Rain Forest*

Ghost forest — where roads are constructed with inadequate drainage, local flooding kills trees (Manaus, Brazil).

is now totally different from the old-fashioned classic view expressed by Jacobs and his translators. For example, the soil below those virgin white-sand (kerangas) forests so far studied is not nutrient poor. The treatment of both macro- and micro-evolution and of pollination biology is heavily biased to the views of a few authors, and does not fairly reflect modern thinking on these important motors of species richness.

The other difficulty, I think, is that the painstaking, reductionist, nuts-and-bolts approach to rain-forest science which has taken us so far in recent years is not compatible with the lyrical style favoured by Jacobs. It should be possible to convey the excitement of science at the same time as conveying its details, but the book which does that for all the world's rain forests has not yet been written.

I foresee endless problems for students who read within this book simplistic glosses on biological complexity, or do not notice the many errors; for example, Kramer's gaps were $1-3 \times 10^3 \text{ m}^2$ not 10^2 in^2 (p.91) or $1-3 \times 10^2 \text{ m}^2$ (p.215). Despite these blemishes this is an outstanding 'first encounter' with tropical rain forests, and a compelling plea — a *cri de coeur* — for the conservation of large tracts of them. □

T.C. Whitmore is at the Oxford Forestry Institute, South Parks Road, Oxford OX1 3RB, UK. His own book, *Tropical Rain Forests of the Far East*, 2nd edn, has just been published in paperback by Oxford University Press.

Looking for a good story

R. W. Guillery

Eye, Brain, and Vision. By David H. Hubel. W.H. Freeman:1988. Pp.240. £14.95.*

OUR ideas about the role that individual nerve cells may play in perceptual experience owe much to relatively few people. David Hubel and Torsten Wiesel contributed significantly to the development of these ideas. In the 1960s and 1970s, working first at Johns Hopkins and then at Harvard, they showed how cells in the visual cortex respond to simple but rigorously specified visual stimuli. They described the characteristics of these stimuli and classified cortical cells accordingly. They went on to show that the arrangement of cortical cells can be analysed in terms of cortical 'columns' and 'hypercolumns', and these came to be seen as basic units of cortical organization in general. In addition, and perhaps most interestingly, they demonstrated that the development of normal response properties in cortical cells depends upon normal visual experience early in life. Early deprivation in one eye, or both, produces severe and irreversible visual losses, which are due to an abnormal development of the pathways that pass from one eye to the cortex.

In *Eye, Brain, and Vision* David Hubel summarizes these studies, setting them in relation to early observations on the retina, the nerve impulse and synaptic transmission. He shows how his own work with Wiesel grew out of the now classical analyses of Hodgkin, Huxley, Eccles, Katz and Kuffler. It makes a good story and is clearly presented for the general reader, "... with a scientific training but not an expert in biology, let alone neurobiology".

As would be expected, the book owes much to the *Scientific American* approach. Problems are stated in as simple and straightforward a manner as possible, and solutions are discovered by people with two names and one place of work (here not always correctly given). The text is laced by Hubel's personal style, which favours jokes and occasionally uses humour to dispose of difficult issues. He avoids complex terms where simple ones can be found, and this is generally, but not always, helpful. 'Nucleus', which is awkward because it has two different meanings in neurobiology, causes trouble. First it appears as 'globes', then 'balls' but finally turns into 'nucleus'.

Most of the book deals with experi-

* In the United States published as volumes in the Scientific American Library.

ments done in the 1960s and 1970s, and could well have been written in the mid-1970s. Not only the visual pathways, but the nerve cells they use belong to this era. Important recent research on multiple visual areas receives only brief mention, and neither parallel 'X', 'Y' and 'W' pathways nor studies demonstrating a role for inhibitory interneurons in the cortex have a part in the book. Discussion of some of Hubel and Livingstone's more recent work on colour-opponent cells in the 'blobs' of the visual cortex is included as a brief look into the 1980s.

There is an important and puzzling departure from general usage in the account of cortical 'modules'. Hubel's suggestion that the modules are larger in cortical layers 3 and 5 than in layer 4 goes entirely counter to current concepts of what a module is. The argument is interesting, deriving from the original defini-

tion of a hypercolumn, and I suspect that, if followed to its logical conclusion, the argument might lead to the (not unwelcome) death of the module.

The book is well illustrated with striking colour photographs and micrographs. It also has some unlikely drawings of nerve cells, which look more like old-fashioned kites flying in the cortex than like neurons. Because they purport to show how simple and complex cells are interconnected, the resemblance may be intended. Finally, there are some marvellously revealing glimpses of Ramón y Cajal playing chess and of the Harvard neurobiologists at play. They provide a clue to the mixture of serious and frivolous that has long given Hubel's style a most readable quality. □

R.W. Guillery is the Professor in the Department of Human Anatomy, University of Oxford, South Parks Road, Oxford OX1 3QX, UK.

Picture chemistry

Murray Goodman

Molecules. By P. W. Atkins. W.H. Freeman: 1987. Pp.197. £14.95.*

IN THE preface to this book, P. W. Atkins describes the beauty and joy he finds in molecular structures. He has selected topics for inclusion which appeal to him and which are evident in the world about us. There is no particular sequence to the chapters in which subjects are covered; rather, Dr Atkins takes a phenomenological view of molecules and their effects. Generally, the molecules are grouped into more or less integrated chapters in which very attractive photographs are coupled to brief accounts of molecular properties.

The book contains descriptions of molecules as simple as diatomic structures and as complex as macromolecules. Should the reader desire, it is possible to find information on structures which elicit taste and smell; which are involved in colour and vision, and smog and air pollution; and which act as drugs and even form materials. The facts are briefly described but the underlying science and interrelationships are omitted. In the introduction, Dr Atkins uses ball and stick representations, assigning colours to denote specific atoms that make up the molecules, and throughout the book the ball and stick structures, space-filling models and structural formulae are presented in a readily comprehensible fashion.

In "Simple Substances", the first chapter, the author describes small molecules that are claimed to be the basis for smog, pollution and acid rain; the relationships presented here are too elementary. In like manner, Chapter 2 contains a systematic treatment of hydrocarbon molecules and

simple alcohols. Suddenly, there is a leap to gamma-aminobutanoic acid, a neurotransmitter. The brief explanations offered for the neurochemical interactions of ethanol, valium and gamma-aminobutanoic acid once again are oversimplified (incidentally, I don't see how one can tell the chirality of lactic acid from crystal birefringence, as is claimed in the caption to the photograph on p.49).

The chapter on synthetic and natural polymers is the best in the book because polymer structures are directly related to their properties. Dr Atkins is, therefore, able to tie the structure of various polymeric systems to their applications as fibres, engineering plastics, elastomers, adhesives and so on. In the last three chapters, a grasp of complex molecular reactions is necessary to understand the relevant effects. But the glossary-like presentations do not provide the necessary insight, and when the author outlines receptor sites such as on pp.44 (protein channels), 107 (taste), 125 (odour) and 165 (morphine), he uses cartoons and structures which are scientifically naive.

What, then, is the value of this book? Those seriously involved in chemistry will find it of little use. But for the chemically uninformed it is an eye-catching introduction to certain molecular effects, and students could be attracted by the photographs and magazine-like brevity of the text. It has often been said scientists do not communicate the content of their work to the public. Chemistry in particular suffers from a poor image. This book could serve the useful purpose of interesting people in the molecular world about us, and thus help to establish chemistry as the central science devoted to molecular studies. □

Murray Goodman is a Professor in the Department of Chemistry, University of California, San Diego, La Jolla, California 92093, USA.

Hard graft in the vineyards

Cyril Ray

The Great Wine Blight. By George Ordish. Sidgwick and Jackson: 1987. Pp.246. £12.95.

WRITING in the 1930s, Winston Churchill recalled that it was in 1898, waiting to go into action on the North West Frontier, that he overcame his "repugnance to the taste of whisky . . . of course all this whisky business was quite a new departure in fashionable England. My father for instance would never have drunk whisky except when shooting on a moor . . . he lived in the age of the 'brandy and soda' ". This merits a footnote in the social history of England, as, in the history of France, do the champagne riots of 1911; then, for nine months of violence and military law, the wine-growers were outnumbered by regular troops, horse and foot.

The social rise of whisky and, thanks to the riots, the existing rigorous laws that govern the production of champagne were indirectly brought about by a bug discovered in a Hammersmith greenhouse in 1863 — *Phylloxera vastatrix*. It ravaged the vineyards of France from then until the turn of the century, costing the country, it is seriously estimated, more than the

Franco-Prussian war.

The phylloxera, which destroys the roots of the European vine, *Vitis vinifera*, swept northwards through France, reaching the Cognac region in about 1875, Champagne in 1890; by the time of which Churchill wrote, British stocks of pre-phylloxera brandy had run low, and were being augmented with spirits made from who knows what, or stuff chemically 'improved'. At the same time it became legal to blend the aromatic single-malt whisky with more nearly neutral grain whisky, to produce a lighter drink suited, with soda, for the drawing-room, leaving single malts for the grouse-moor.

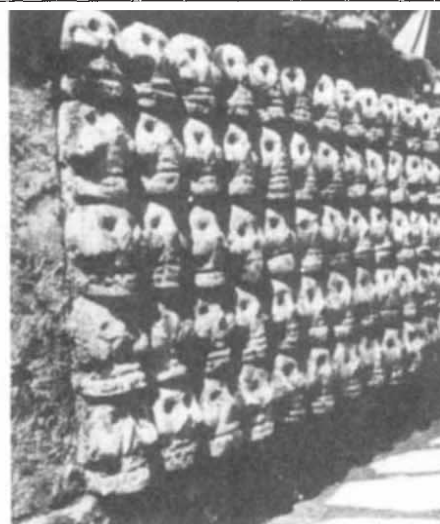
George Ordish was not only producing wine in Kent 50 years ago but for at least as long has been one of Britain's greatest authorities on pest control. In this book, first published in 1972, he does not neglect such social consequences as this of the phylloxera plague. But in the main he recounts in scholarly fashion the many chemical and other remedies that were tried and found wanting for half a century or so from the 1860s, until it was shown that the classic vines of Europe, producing fine wines, would go on doing so after being grafted on to the root-stock of native American vines, producing wines that were far from fine but were resistant to the phylloxera.

So far, so good. The last chapter of this revised edition, however, brings the dread news that a new, sturdy strain of

phylloxera has appeared in many parts of the world — even in the United States, whence had come what had been thought to be salvation.

Tiny vineyards exist where ungrafted vines still flourish. In Champagne, the distinguished house of Bollinger has two half-hectare plots where ancient vines are 'layered' to send out roots each spring from buried branches. Since 1969, at my suggestion, Bollinger have made wine from these vines separately from the others, but Mr Ordish will be disappointed in the hope (expressed on p.137) that this would make possible a comparison between wines from ungrafted and from own-roots vines but of the same varieties of grape from the same soil. Not so: since 1974 the pre-phylloxera wines, in exceptional years, and at necessarily high prices, have been sold as Bollinger Vieilles Vignes Françaises, and show themselves deeper in colour and in flavour than Bollinger's other wines. But this could well be, if only partly, because the ancient vines are all Pinot Noir, providing that champagne rarity, a *blanc de noirs*, whereas all other champagnes are a blend of black grapes and white — mostly black — or the lighter *blanc de blancs*. □

Cyril Ray, Albany, Piccadilly, London W1V 9RQ, UK, formerly a war correspondent and foreign correspondent, is author of Bollinger: The Story of a Champagne (1971) and The Wines of France (1976) among many other books on wine.



Primal scream — a stone head of Xipe Totec, the flayed god associated with vegetation, wearing his flayed skin. And (above) a view of a sculpted skull-rack facade. The pictures are taken from *The Great Temple of the Aztecs: Treasures of Tenochtitlan*, an account by Eduardo Matos Moctezuma of the rediscovery of the temple in 1978. The book gives a detailed account of the excavation itself, along with an analysis of the development of the temple through six building phases, and places the discoveries in historical context. Publisher is Thames & Hudson/W. W. Norton, price is £24, \$29.95.

Beating the plague

Len Goodwin

Smallpox and its Eradication. By F. Fenner, D. A. Henderson, I. Arita, Z. Ježek and I. D. Ladnyi. *World Health Organization: 1988. Pp.1,460. SwFr. 250, \$150, £80.*

SMALLPOX was with us for a very long time — probably ever since the development of irrigated agriculture 10,000 years ago, which initiated the human population explosion that still continues today. The variola virus may have already existed among our primate ancestors, or have been derived from another animal orthopox virus, but it certainly got its opportunity when large numbers of people began to live in close contact. It had no alternative animal host, no insect vector and it produced lasting immunity: it could therefore only persist if enough susceptible human beings were available to maintain a continuous chain of infection. Smallpox was universally feared and, in the cities of Europe and Asia where records were kept, caused on the average 10 per cent of all deaths every year. These characteristics, together with an easily recognizable rash and the existence of a safe and effective method of immunization, made it possible to plan a world-wide campaign of eradication.

The elimination of smallpox was first mooted to the World Health Assembly in 1953 but the idea was turned down. Five years later a global eradication plan was accepted and put into operation. But although the disease was removed from some 30 countries, inadequate funding meant that the hard core in the Indian subcontinent and in Africa remained. In 1966 an intensified programme with better financial backup was instituted, and almost achieved its ambitious target of eradicating smallpox within ten years. In fact it took eleven; despite intensive surveillance, no endemic smallpox has been detected anywhere in the world since the last case in Somalia in October 1977.

Edward Jenner was largely responsible for the introduction of 'vaccination' with cowpox virus to replace 'variolation', a method that used material from the pustules or scabs of smallpox patients and carried a risk of up to 2 per cent mortality. In 1801 Jenner wrote: "the annihilation of the Small Pox, the most dreadful scourge of the human species, must be the final result of this practice". The fulfilment of his prophecy, 176 years later, represents a triumph for the World Health Organization and is meticulously celebrated in this great, big, beautiful book that took seven years to write.

The style is clear and interesting, there are hundreds of excellent colour plates,

maps, tables, photographs of places and people, and plentiful 'boxes' containing summaries of the text. It is a worthy tribute to the work of Donald Henderson and his team. Everything is covered — early history, basic science, clinical aspects, pathology and immunology, vaccine production and application. There are detailed accounts of the complex operations, different in every country, which were developed for the effective deployment of thousands of local staff for the detection of cases, vaccination, follow-up and surveillance, often in remote and sometimes dangerous terrain. Vaccine was improved and stabilized, and

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

This was smallpox.

its administration was facilitated by the introduction of a bifurcated needle that picked up the correct dose for inoculation. House to house searches for cases were made, in which people were shown a picture of a child with smallpox and were asked whether anyone had seen a person with such a rash. Helicopters were used to reach remote valleys in mountainous country. All of these operations were complicated by a host of natural and political problems — floods, drought, famine, war, incompetence, dishonesty, personal antagonisms, shortage of local resources and lack of cooperation.

Success is attributed to four main factors: a central international organization which had the ability to express policy directly and firmly to governments; a

dedicated and competent national and international staff, willing and able to adapt and to innovate; a plan based on principles and methodologies, *not* directives, so that operations in each country were different, and changed appropriately as the campaign progressed; and appreciation of the need for continuing field and laboratory research, and the availability of strong international technical support (which was freely given).

Eradication is a bold word. What are the chances of the reappearance of smallpox? After ten years, the virus is unlikely to have remained undetected in the field, either in man or in any other primate that, on close contact with man, has been shown to be susceptible. Monkeypox caused concern when it appeared in West and Central Africa during the 1970s, but DNA analysis showed that it differs from variola virus, for which there appears to be no animal reservoir. In China, Afghanistan, Pakistan and Ethiopia, where variolation was widely practised by traditional experts, any remaining reserves of scabs from smallpox lesions are unlikely still to be infective and the chance of virus surviving in corpses, even those buried in the permafrost, is regarded as remote. Laboratory stocks of variola virus clearly constitute a risk, as shown by the outbreak in a Birmingham laboratory in 1978; strenuous efforts have been made to destroy all the remaining samples lurking in laboratory deep-freezers. The only reserves of virus are kept for reference in two maximum-security freezers, one in Moscow and one in Atlanta, Georgia.

There remains the possibility of deliberate release, although smallpox, because of its slow spread by face-to-face contact, is not an ideal biological weapon. Nevertheless, there is always a chance that some idiot might one day use the virus to threaten an unvaccinated world. As the authors conclude:

it is melancholy but perhaps realistic to suggest that the possibilities of biological warfare or terrorism now constitute the chief reason for holding reserves of vaccine and for maintaining epidemiological and laboratory expertise for the diagnosis and control of smallpox.

In a world of nuclear weapons and other menaces it is good to see that science at its best, backed by effective organization, can really work to the benefit of mankind. Arthropod-borne diseases with reservoir hosts, and those that depress or fail to stimulate the immune response, will be less easy to conquer. But the pattern of success has been established and WHO should face the future with renewed confidence. □

Len Goodwin, Wellcome Trust Film Unit, 183 Euston Road, London NW1 2BP, UK, was formerly Director of the Wellcome Laboratories of Tropical Medicine and Director of Science at the Zoological Society of London.

The drawbacks of dynamism

Robert Olby

Controlling Life: Jacques Loeb and the Engineering Ideal in Biology. By Philip J. Pauly. Oxford University Press: 1987. Pp.252. \$24.95, £22.50.

PROFESSOR Pauly's study of the life of the physiologist Jacques Loeb (1859–1924) and of his influence in the United States is a remarkable *tour de force* which opens up a fresh chapter in the history of modern biology. The style is lively, the organization clear, and the impressive support from documents is successfully compressed within the notes so as to relieve the pressure on the main text. My only reservation about the book's form is the absence of a bibliography of Loeb's writings and of a separate bibliography of the sources cited in the notes.

Pauly takes us through his subject's career, which began in Germany in the 1880s. From the start Loeb's enthusiasm for physiological research brought him into conflict with the authorities of the German physiological community. He sided with his teacher, Friedrich Goltz, in opposing the 'modern phrenology' of the localizers, Munk and Hitzig. Loeb argued for the integrity of the organism, and throughout his working life he preached the holistic approach. Failing to find a niche in animal physiology, Loeb turned from the constituency of medicine to Julius Sachs, the eminent plant physiologist, whose institutional links had been with agriculture. The young Loeb found Sachs's work on tropisms exciting for the potential it offered to account for aspects of animal behaviour — in particular spatial orientation — which he wanted to explain in a non-dualistic manner. From this time on, tropisms played an important part in Loeb's research, and he was soon demonstrating 'tropistic' responses in insects and crabs.

Meanwhile Loeb began to correspond with Ernst Mach, the physicist and positivist. He admired Mach's negative attitude to causes and mechanistic explanations. Mach's rejection of the separation of science from technology, and his appreciation of scientific concepts as tools for achieving control and prediction of nature resonated with the views of Loeb's teachers, Goltz and Sachs. Thus crystallized in Loeb's mind the view of science as the means to control nature, giving man the power to fashion it to suit his needs. Nature undisturbed by man was one state

among many possible states which, by invention, the scientist could bring into being. For Loeb, the former had no privileged status. Consequently, the question of how evolution had taken place had no particular appeal to him, whereas how scientific knowledge could be used to change species did. Success in the second enterprise might well precede discovery of the mechanism of the internal workings of the organism. Such a discovery involved inventing mechanisms as causes, an unattractive task to a Machian positivist.

Loeb's marriage in 1890 to Anna Leonard made the need for a satisfactory academic position pressing and in 1891 he became an associate professor at Bryn Mawr, as "no other suitable candidate had been presented". The invitation to join Charles Otis Whitman at the new Univer-

provide "complete reductionist explanations" of the biological phenomena which hitherto he had been content to seek only to control. The result was that he launched into a seven-year study of protein chemistry and retreated from his earlier commitment to a broad programme of research in experimental biology.

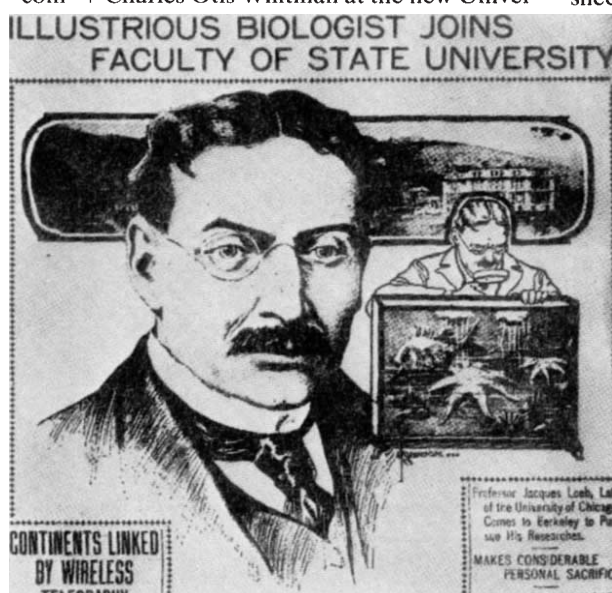
Pauly considers that this change of direction caused a "substantial decline in [Loeb's] scientific creativity", and he lays the blame for this on his "increasing emphasis on reductionist explanations" and the pressure he felt to make good as a scientific genius. Pauly also concludes that the image of Loeb that we get from his famous address of 1911, "The Mechanistic Conception of Life", is of the man who had become a physical chemist having shed his earlier role. This image is there-

fore unrepresentative of Loeb. To stress the point the author details the influence of Loeb upon such important figures as J.B. Watson, the founder of behaviourism; B.F. Skinner, Watson's most prominent successor; the geneticist, Muller; the physiologist, Crozier; and the inventor of the contraceptive pill, Gregory Pincus. These men were committed to the creation of technology in the service of man — Watson and Skinner to behaviour therapy, Muller to mutagenesis and Pincus to contraception.

This perspective on Loeb and his American admirers throws much light on their aims and achievements, their places in the scientific community, and the constraining influences of their institutional affiliations. It also raises problems: if Loeb had suffered

such a change of direction in his scientific career, why do we find so consistent a commitment in his work to physico-chemical explanations? What of other possible factors which might have influenced his late entry into the physical chemistry of proteins? Consider F.G. Donnan's ingenious theory of membrane equilibria (1910), Procter and Wilson's successful application of the Donnan equilibrium to gelatin (1916) and the introduction of methods to measure hydrogen ion concentration (1910). (It is curious that Pauly does not even mention Donnan, Procter and Wilson.) Surely the biologist who induced parthenogenesis by changing the salt concentration of the liquid surrounding the egg had good reasons for exploring the effects of changes in acidity on a protein and the effects of salts on protein colloids?

Pauly's image of Loeb's biological research as engineering is powerfully presented. It goes a long way towards explaining why Loeb's philosophy of biology, and indeed of life, seems so



Front page news — the San Francisco Examiner announces Loeb's move to Berkeley in 1902.

sity of Chicago followed in 1892. For the eleven years that Loeb worked in Chicago he was at the heart of the movement known as American progressivism. Loeb shared this movement's aim to use science to transform society, but he strongly opposed the evolutionary assumptions upon which it was built. Evolution, he believed, would be established by experimentation, not by speculative, descriptive phylogenies, and imaginary architectures of the germ-plasm. Accordingly he applauded De Vries but not Weismann, and devoted himself to the experimental study of fertilization, parthenogenesis, tropisms and embryological development, a programme which he was able to pursue single-mindedly when the University of Berkeley made him a research professor.

Pauly detects a profound change in Loeb's epistemological views in 1910 which had consequences for his research work. This was also the year in which he joined the staff of the Rockefeller Institute. In this "Palace of Pure Science", writes Pauly, Loeb at last felt compelled to

From *Controlling Life*

shallow and unsophisticated. Whether Loeb's scientific career should be interpreted in terms of the contrasting reductionist and anti-reductionist research strategies is, however, open to question. To be opposed to localization theories of the functioning of the cerebral cortex does not entail opposition to physico-chemical explanations of brain processes. Loeb's holism was not at odds with his reductionism — we just need to be careful what we mean when we apply the term reductionist to him.

Two contrasting approaches which are surely more appropriate to his work are those of dynamical and structural explanations. Among the former are vibrations, patterns, gradients, fields (all of them dynamic organizations due to continuing physical processes), whereas structural explanations refer to persistent morphological entities (chromosomes, nucleoli, cytoplasmic particles) whose existence

does not depend on continuing processes. Loeb, like d'Arcy Wentworth Thomson and William Bateson, relied on dynamical explanations. Loeb turned to colloidal systems and to the Donnan equilibrium, Thomson and Bateson to analogies with magnetic fields, sound waves, sand ripples, and the stresses and strains of bridges. All three were opposed to Weismannian preformation and the artefacts of the cytologists. Yet it was the successes of the cytologists and cytogeneticists which justified the identification of morphological structures in the cell as causal agents of its behaviour. No wonder Loeb, like Bateson, seemed to have lost his way in biology and only rescued his reputation by his study of the physical chemistry of the proteins. □

Robert Olby is Reader in the Division of History and Philosophy of Science, Department of Philosophy, University of Leeds, Leeds LS2 9JT, UK.

altogether different pattern that he had chosen to follow from the start.

Had he not had to work within these preordained bounds, Stearn would surely not have arrived at the conclusion, as he does in his preface, wistfully and half-apologetically, that "on the whole the Society's life has seemed unexciting, almost at times lethargic". For there is much in the surviving records from which a livelier tale could have been extracted by a pen more agile and experienced than Gage's. Even the second quarter of the nineteenth century — when the society had beggared itself in purchasing Linnaeus's collections and had fallen under the sway of a deadeningly unimaginative group of elders — was the occasion of many a colourful shaft in the published and unpublished literature which it would have been helpful to have disinterred. The personality of the founder, James Edward Smith, needs a less sparing analysis than it has received as yet, and there is reason to believe that the circumstances of the society's birth were considerably more complicated than the received account suggests.

The famous meeting of 1 July 1858, at which the papers of Darwin and Wallace were read jointly (in the absence of their authors), is well described and the reasons for their failure to make an impact are examined in justifiable detail; at the same time it is hard to believe that the Darwin industry has not now uncovered a greater range of letters to enable the affair to be documented more thoroughly. Gage was good, too, on the long and contorted struggle that had to take place before women were admitted as Fellows, but his account ought to have been amplified by the more recent research which would have enabled it to be seen within a wider context. In the same way, more comparisons with the experience of the Linnean's sister societies in London, in particular its offshoots, the Zoological and the Royal Entomological, would have been enlightening: Gage's view was too myopically domestic.

All the same there is an enormous mass of information here that historians of science will be eager to quarry. The Linnean Society has played such a central part on the natural-history-cum-biological stage over such a lengthy span of years that no student of the development of biology or scientific institutions can afford to bypass its extraordinarily comprehensive records — or the many collections of letters that so richly augment them. In preparation for their visits, historians will assuredly find this volume to be an indispensable introduction. □

David E. Allen is in the Unit of History of Medicine, University College London, Gower Street, London WC1E 6BT, UK, and author of The Botanists: a History of the Botanical Society of the British Isles through 150 Years (St Paul's Bibliographies, 1986).

High society

David E. Allen

A Bicentenary History of the Linnean Society of London. By Andrew Thomas Gage and William Thomas Stearn. *Academic*: 1988. Pp.241. £25, \$49.50.

"THE world's oldest active biological society", the Linnean ranks in seniority second only to the Royal in that pantheon of learned bodies, steeped in history and majestically housed, which for so long constituted 'Scientific London'. Through their meetings and their journals, the two societies continue to play a leading role not only in the metropolis but far beyond. The Linnean, however, has the extra and unique function of acting as international caretaker of the collections of type specimens which were the original pretext for its founding, and on which the whole huge edifice of biological nomenclature ultimately rests. The rooms in Burlington House, the society's home since 1873, consequently double up as what is effectively a shrine.

Such a body can hardly help but be proudly conscious of its past, and it was unthinkable that the series of events with which it is celebrating its bicentenary should not have included a volume updating the story of its life. There was a problem, however, in deciding how best his should be done. For, awkwardly, the first century and a half were thoroughly chronicled in 1938 by A.T. Gage, one of the society's former executive officers. Though clogged with too much loving detail, and not very enthrallingly written, Gage's history was too substantial to be disregarded lightly.

To have gone over that ground all



'To learn the ways of nature', the Linnean's motto. afresh might well have amounted to little more than a work of duplication; on the other hand, to have produced a book limited to the latest half-century would have meant consigning Gage's efforts — and the greater part of the society's collective memory — to almost certain oblivion. In the event, Professor Stearn, a recent past president, the leading authority on Linnaeus and author of a masterly and most readable account of the development of the British Museum (Natural History), has settled for the compromise of a rounded-out 'Gage', adding five chapters to the narrative and taking down to the present the various lists, statistics and sections on particular activities which occupied a third of his predecessor's text.

Hard though that decision must have been, many of his readers will nevertheless regret that Stearn shied away from the larger task of tackling the subject *de novo* and of providing the history that he was so supremely well-fitted to produce. As it is, he has condemned himself to falling in with Gage's style and treatment as faithfully and as self-stintingly as feasible; and although he is to be congratulated on a difficult feat of knitting which has resulted in a fabric without noticeable seams, one cannot help wishing that it had been an

The terrier's way

Ryuichi Inoue and Alun Anderson

Imanishi Shinkaron Hihan no Tabi.* By Beverly Halstead. Translated by Akiko Nakayama and Suisen Sakuramachi. *Tsukiji Shokan, Tsukiji 2-10-12, Chuo-ku, Tokyo 104, Japan:1988. Pp.282. Pbk ¥2,200.*

BEVERLY Halstead, Reader in the Departments of Geology and Zoology at the University of Reading in Britain, was once described by a colleague as "Darwin's terrier". Terriers are small dogs, once used to hunt rats. Nevertheless, Halstead says that the description appeals to him. He admires the terrier's sharp teeth and its unwillingness to let go. Like T.H. Huxley, "Darwin's bulldog" in the nineteenth century, Halstead has been vigorous in his defence of mainstream evolutionary theory. Most memorable was his attack on the cladists in 1980, which triggered a debate that occupied the correspondence columns of *Nature* for a year.

In his new book, Halstead tells how he scented anti-Darwinism while on a short visit to Kyoto University in 1984. Ostensibly, the book is an account of his pursuit of the theories of Kinji Imanishi, Emeritus Professor of Zoology at Kyoto University, now over 80 years old and a prolific author of great reputation in Japan. But the book is an unusual one; it is published only in Japanese and it fits into no simple category. It is part travelogue, part science text, part personal diary and part exposé.

With little time in Japan, Halstead attacked his subject with manic energy. It mattered little that all of Imanishi's works are in Japanese; he persuaded students to paraphrase them into English. It mattered less that many of the people he wished to interrogate could not speak English; he press-ganged anyone he could find into taking on day-long interpreting sessions.

Halstead's intellectual progress provides the best part of the book. He reveals not just his conclusions — an account of Imanishi's theory and its faults (see *Nature* 317, 587; 1985) — but also his ways of digging and delving, his blunders and his successes.

Finally, Halstead met Imanishi. They agreed about almost nothing, but Imanishi gave him his seal of approval. Halstead, it seems, had understood his ideas. Halstead, however, wished to go further. He wanted to understand the whole Imanishi phenomenon; why it is that an evolutionary theory which appears to defy common sense by denying both intra- and inter-specific competition has become popular in Japan. He looked for

answers in Japanese society, and for the Japanese reader the story of his search will be a delight. Each new experience sent the terrier off in a wholly unexpected new direction: he met the great and the eccentric in Japan's geological community and discovered phenomena — such as the Chidanken's 'democratic geology' and mass excavations — that few foreigners ever see. On this journey he dug deep into people's backgrounds, and in the book he speculates fiercely about Japan and his Japanese friends.

To understand how this seems to a Japanese, the Western reader might imagine the translated account of a Japanese professor, who speaks no English, who has spent a few months at Oxford and who gradually works out, through the characters he meets, Oxford's contradictions and Britain's hidden class structure.

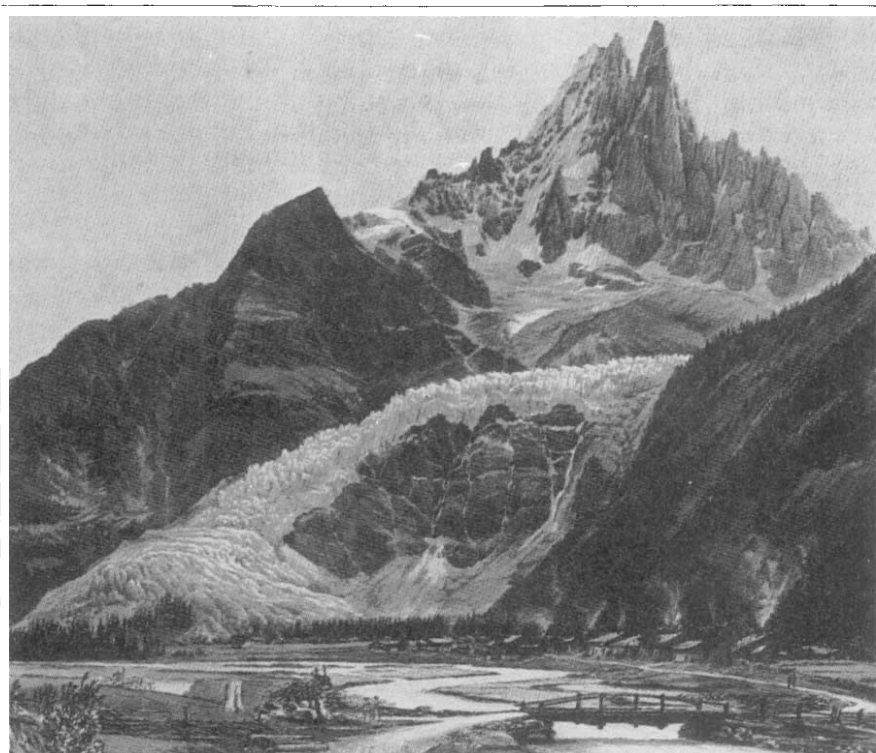
Ultimately, however, Halstead falls into a trap. He defies the old maxim that "to write a book about Japan you must stay three weeks or thirty years". He stayed a little too long, and in the book he passes from fresh first impressions to oversimplistic theories about the nature of Japan. His final view, that Imanishi's theory has been popular because it offers a dream world of harmony as an antidote to the cruel reality of modern Japan, is naive.

Even though Halstead may be wrong on this count, it matters but little. The path that led him to that view is frankly charted



and his mistakes are honest ones. More than for its attack on the Imanishi phenomenon, the book will be admired in Japan for a very un-Japanese quality: its brave refusal to compromise. The terrier's message is that everything can be questioned, and his way of action is on open display. □

Ryuichi Inoue, now retired, was formerly a biology teacher in Kanazawa, Japan. Alun Anderson is Washington editor of *Nature*.



Coming down the mountain — in 1820 the Glacier des Bois, the tongue of the Mer de Glace in the Alps, "attained a distance of only 65 yards from the house of John Marie Tournier, the nearest in the village, where its further progress was providentially stayed" (J.D. Forbes). This 1823 painting of the glacier is taken from *The Little Ice Age* by Jean M. Grove, to be published next week by Methuen, price £85.

*Imanishi's *Theory of Evolution: A Critical Journey*.

But does it please the Prince?

Daniel S. Greenberg

Science and Technology Advice to the President, Congress, and Judiciary. Edited by William T. Golden. Pergamon: 1988. Pp.523. Hbk £51, \$67; pbk £9.25, \$13.25.

IN RECENT years, the leaders of American science have mainly been snubbed in their devoted attempts to provide presidential science advice as they believe it should be provided. You need us, whether or not you realize it, they have earnestly told a succession of presidents. But presidents tend to see the matter as Machiavelli suggested: "A Prince . . . ought always to take counsel, but at such times and seasons only as he himself pleases, and not when it pleases others".

Convinced that his science advisers were politically undisciplined tattlers (as some indeed were in opposing the supersonic transport vehicle and the anti-ballistic missile), Richard Nixon petulantly abolished the White House Office of Science and Technology and its closely affiliated council of wise men, the President's Science Advisory Committee (PSAC). Restored late in the Ford administration at the behest of Congress, the science office held only limited authority under Jimmy Carter. Long afterwards, the committee was revived in skeleton form only. For the first five months of the Reagan administration, the science office existed on a caretaker basis until the appointment of a little-known government physicist, George A. Keyworth II, as director. When he left at the end of 1985, the directorship remained vacant for ten months while the staff dwindled away. Several well-regarded elders of science declined invitations, feeling the job had eroded to a near meaningless state. The eventual appointee, William R. Graham, a right-wing nuclear weapons researcher on the rebound from a misbegotten appointment at the helm of NASA, entered the office with a unique distinction: denunciation by several senators as unqualified for the job. During his brief tenure, the office has sunk so low that the condition of presidential science advice was the subject of an unprecedented congressional hearing in February. Science and government at the White House thus fall short of the noble intimacy prescribed by C.P. Snow and his spiritual brethren.

Does it matter? Obviously not to recent presidents, nor, if campaign statements are the measure, to the current crop of presidential candidates. Nonetheless, the would-be courtiers press on, led by the patron saint of White House science advice, William T. Golden, a man about Wall Street and science and government for nearly 40 years. Asked by President

Truman to recommend a structure for presidential science advice, Golden, in 1950, proposed an in-house presidential science adviser working in close association with a presidentially appointed part-time committee of distinguished seniors drawn from the scientific enterprise. Presidential interest in this proposition may be inferred from the fact that it wasn't until 1958, the Eisenhower presidency and the political trauma of Sputnik that the Golden prescription was implemented.



Executive home, with William Golden, "the patron saint of White House science advice".

Thus began what is nostalgically regarded by many of its participants as the great period of presidential science advice, an era in which Killian from MIT, DuBridge from Caltech and Kistiakowsky from Harvard were at Eisenhower's elbow, while similarly distinguished colleagues from academe routinely commuted to Washington to advise the adviser and even meet the president. The presumably glowing relationship continued into the Kennedy administration, with Wiesner of MIT as the president's science adviser. From arms control to science in the schools, from energy to international research cooperation, nothing was beyond the purview of these high-riding scientist-statesmen, especially the care and feeding of science in the universities, whence virtually all of them came. To outsiders, these people looked very much like an in-house lobby for their own profession. But with heartfelt insistence, many of the survivors of this era claim to have performed invaluable, unique public service.

William Golden, contending that "our country has been the poorer" for their

dismissal from court, yearns for a restoration. In 1980, he commissioned a collection of brief articles, published as *Science Advice to the President*, arguing the old theme, that you need us, whether or not you know it. Now, returning to pursuit of that goal, he has assembled this unruly *mélange* of 85 commentaries, mostly focused on science advice for the presidency, with a few devoted to Congress and the judiciary.

About a dozen or so of the essays, mostly by veterans of the presidential advisory service, are firmly rooted in the political realities of life in the White House; they are useful contributions to dealing with the persistently difficult problem of relations between political power and specialized knowledge. Other articles can be characterized as the deliria of the naive and innocent. One writer, for example, urges the president to "sponsor

an annual, heavily publicized prize competition for highly promising, practical, strategic new thrusts in science". And another advises Mr Reagan to consult Mrs Thatcher if he wishes to emulate "the resurgence of British science, technology, and industrial growth in the 1980s".

Perhaps the sagest contribution is provided by one of the few wise men and original thinkers of American science policy, Alvin M. Weinberg, former director of the Oak Ridge National Laboratory and a member of PSAC during the golden era, 1958–1961. Weinberg usefully deflates the myth of PSAC's omniscience and influence by noting that "On the decision to send a man to the moon, PSAC was ignored . . . largely, I believe, because most of its members were against the mission". Weinberg correctly observes that presidential science advice was "unorganized and sporadic" when Sputnik inspired Eisenhower to install science advice in the presidency. Today, he points out, "science advice pervades our governmental agencies and departments. Such advice is, above all, institutionalized

and bureaucratic". And he offers an intriguing proposition: that the Strategic Defense Initiative, widely opposed by scientists as militarily provocative and technologically not feasible, reflects sound political instincts that transcend scientific considerations. Weinberg concludes that "[this administration's] decisions on science matters and on broader policy having scientific components . . . have not been notably worse than in previous administrations where the science advisor had more bureaucratic status. . . ."

On the other hand, Frank Press, president of the National Academy of Sciences and science adviser to President Carter, expresses hope that the "younger breed of presidential candidates" will see that they need science advice close at hand "because of the way society is going". He proposes elevating the job in the White House hierarchy, perhaps to cabinet status without portfolio. And he wants to resurrect PSAC. But Press concedes the ultimate political reality: "You can't force it on the president".

The same point is stressed by Solomon Buchsbaum, of AT&T Bell Laboratories, who has served throughout the Reagan administration as chairman of the White House Science Council, the slimmed-down descendant of PSAC. Noting that "There is much moaning these days about the absence of a 'truly presidential' science advisory mechanism, that is, the absence of a PSAC", Buchsbaum petulantly observes, "the critics expect miracles".

Many who are sceptical about the value of science advice in the White House point out that, in Washington today, science advice, solicited and otherwise, gushes from innumerable organizations, including the National Academy of Sciences, the Congressional Office of Technology Assessment, professional societies, congressional witnesses and so forth. H. Guyford Stever, who served as President Ford's science adviser, estimates that the science advising industry has a "collective budget of a good fraction of a billion dollars", and adds, "I do not see that there are important ideas in science and engineering that do not get the attention of the top decision-makers in our country".

The strongest case for the importance of presidential science advice is offered in the volume's only contribution by a recipient of such advice, Gerald Ford, who signed into law the restoration of the White House science office. In justification, he explains: "Presidents don't like surprises". That may be the strongest argument for putting scientific skill directly in the service of the presidency. □

Daniel S. Greenberg, 3736 Kanawha Street NW, Washington, DC 20015, USA, is editor and publisher of the newsletter Science & Government Report.

Assessment of the assessors

Philip Gummert

Evaluating Applied Research: Lessons from Japan. By John Irvine. *Frances Pinter/Columbia University Press: 1988. Pp. 103. £27.50, \$45.*

THE word 'evaluation' has come a long way from its origins as a term used in the insurance business of the eighteenth century. It now increasingly refers to the latest fashion in science policy, namely, the attempt by governments and other bodies to appraise the quality and benefits of scientific and technological research and development. The search is on for quick, cheap ways of making such appraisals, preferably through the development of sets of 'performance indicators' such as numbers of papers published and subsequent citations, patents awarded or some measure of rate of return on the initial investment. As these approaches spread throughout government, industry and the universities, more and more scientists are becoming aware of them, and of the fact that speed and cheapness in their use can be attained only at the price of sensitivity in their application.

Much of the early work concentrated on evaluation of basic research, mainly using bibliometric techniques. Attention has moved more recently to applied research, defined in this book as "original investigation undertaken in order to acquire new knowledge which is directed towards practical aims or objectives". This definition includes the concept of 'strategic' research, where there are likely to be practical applications but they cannot yet be clearly specified. Obviously, counts of papers and citations will be less valuable for evaluating applied research than for basic, and assessment of the value of the work for eventual practical application will be more important.

The shift in attention towards applied research has been a notable feature of Japanese work on evaluation. This is partly because of the Nakasone government's concern to get better value for public money, but it also owes much to the changing circumstances of Japan's science and technology. The rapid expansion of government-funded research of the 1970s has given way to more modest growth. After catching up with the West in many aspects of technology, Japan now feels obliged to pioneer key areas of science and technology. Government laboratories are becoming less concerned with specific applied research in support of industrial ventures, and instead are concentrating more on longer-term strategic research. Finally, there is a growing awareness that

the traditional research management procedures, based on the search for consensus, may not be appropriate where relatively radical decisions have to be made, such as closing an institute.

All this has created a need for a better basis for making today's tough decisions over resource allocation, and for ways of minimizing risks as the emphasis shifts from an almost exclusive concern with product and process development towards the less-charted upstream areas of applied research. Accordingly, systematic external reviews have, for the first time, been introduced to Japanese government laboratories and a number of reports have been produced offering guidelines for research evaluation.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Pointing forward — Japan's Maglev train.

John Irvine offers an overview of Japanese practice in these respects. *Evaluating Applied Research* is basically a report prepared for the British government's Department of Trade and Industry. It is therefore written in report style, complete with "Executive" summary (the superfluous adjective is such a ubiquitous feature of report-writing that it too must be seen as a sign of the times). The book emerged from a five-week study, involving interviews at over 20 Japanese organizations. The main substantive chapters are devoted to the evaluation procedures of the Administrative Inspection Bureau (the specialist evaluation arm of the Management and Coordination Agency of the Prime Minister's Office), the Agency of Industrial Science and Technology in the Ministry of International Trade and Industry, and the Science and Technology Agency. It is a pity that the Defence Agency was not included because, even though still a small spender on research and development by Japanese standards, it is growing very fast, particularly in the more upstream areas.

The systematic evaluation of the 140 government laboratories and national

research institutes is being carried out by a division of 14 staff within the Administrative Inspection Bureau. At eight to ten evaluations per year, it is clearly going to take them some time to complete the first round of appraisals. Interestingly, there is no requirement for the inspectors to have a science degree. They operate by collecting written material from the institute, and then conducting a site visit, which is followed by hearings with external experts or users of the institute. It is already clear that they are willing to recommend closure of programmes of research.

As well as the infrequent attention of the inspectors, institutes and programmes are increasingly subject to review by their parent agency. The Japanese have recognized the need to build up the skills needed for this task; interestingly, they differ from Western countries in making little use of external evaluators from consultancy firms or research institutes. They also differ from at least some Western countries in having no hang-ups about the role of the state in the support of industrially relevant research and development. Projects in the Agency of Industrial Science and Technology have a planned duration of ten years; they are chosen in part because they are expected to have an industrial impact yet cannot be undertaken by private firms because of the risk involved; and the Japanese do not, as in Britain, apply an 'additionality' test (which constrains the state to fund only work that would not otherwise have been undertaken by firms), but instead use government funds to coordinate and target research of the highest priority that would in any case be carried out by companies.

Increasingly, Japanese projects have evaluation criteria built into them from the outset. In this respect they follow standard planning practice of identifying goals and specifying how one will tell whether they are being achieved. Considerable work goes into *ex ante* evaluation of the potential importance of candidate research programmes. Special efforts have been made in recent years to introduce effective mid-term reviews, which enable programmes to be changed and resources concentrated on the more promising areas in the light of experience gained earlier on. Quantitative performance indicators are used, particularly patents and publications, and Irvine detected some evidence that the Japanese government is tempted by evaluation techniques that lend themselves to simple quantification. However, calculations of rates of return are attempted only rarely for government-funded programmes. Instead, emphasis is put upon whether or not projects are achieving previously defined technical and other strategic goals (which might include building a national capability in a particular area of technology). Care is always taken to discuss

the results of the appraisal with the appraisees, not least because of a desire to generate a commitment among researchers themselves to the idea that assessment will not only improve efficiency but is also in their interests.

There is a danger that the evaluator may inherit the mantle of Oscar Wilde's cynic, who knew the price of everything and the value of nothing. To gain the benefits of evaluation while avoiding the pitfalls is a task of growing importance, to which this book makes a valuable contribution. □

Philip Gummert is Senior Lecturer in the Department of Science and Technology Policy, University of Manchester, Manchester M13 9PL, UK.

Rhine maidens

Steve Blinkhorn

Forbidden Knowledge: The Paranormal Paradox. By Bob Couttie. Lutterworth: 1988. Pp.155. £9.95.

THE lunar branch of marine biology is not a thriving field of research. Before 1969 it might have been possible to hope for a renaissance, but it is to the eternal credit of science that no one any more thinks that there are mermaids on the Moon — not even very *small* mermaids.

Sadly, the lesson has not been learned by parapsychologists. In the nineteenth century, belief in a parallel reality, a different order of causality, led to reports of paranormal manifestations on a scale and intensity which demanded attention. Tables flying through the air unaided by forces known to the heirs of Newton; ectoplasm in grocers' quantities; thought transference, spirit manifestations and hauntings of an intensity commensurate with the existence of an independent *élan vital*.



Fairy tales — pictures of the Cottingley fairies, published in Strand Magazine in 1917, caused a storm of controversy. Only many years later did the girls admit the pictures were faked. The picture is taken from The Reality of the Paranormal by Arthur Ellison, published by Harrap price £12.95.

Curiously, the paranormal has become decidedly less spectacular. At best these days we have the spoon-bending antics of Yuri Geller, and laboratory effects which strain for statistical significance by dint of mind-numbing repetition of trials. Parapsychology has retreated to claiming, in effect, that on the Moon are mermaids of microscopic proportions.

Forbidden Knowledge examines the current state of parapsychology from the point of view of an experienced practitioner of long-distance magic. Here is a book that deserved to have been written at greater length. As it is, it effectively dishes parapsychology, to the extent of providing a cookbook of methods for replicating spoon-bending and spirit-communicating stunts. But Couttie's prose style stumbles over itself in haste, and the argument suffers from superficial reference to research rather than detailed discussion. Having made a special study of the field, the author's talent for communicating his knowledge is stretched.

So, as a book *Forbidden Knowledge* fails to satisfy. Nonetheless, it is welcome. The fundamental flaw in popular accounts favourable to the paranormal is accurately identified: only positive results find ready publication. One could add that selective publication subverts all classical hypothesis-testing on a statistical model and that even the best-designed experiments can admit of more than one alternative hypothesis given rejection of the null hypothesis. Hypotheses involving extra-sensory perception, thought transference and the like are difficult to frame in the first place, and the usual statement that a set of results could only have happened by chance at some extreme level of improbability glides over the range of uncontrolled factors which could provide a mundane explanation. Close inspection typically reveals just such uncontrolled factors, and Couttie neatly disposes of a number of highly publicized instances of paranormal

phenomena which were greeted with breathless enthusiasm, but turn out to have been examples of the conjuror's art. So much for the impressive significance levels frequently cited in ESP laboratory reports.

However, if you want to find out how to bend spoons *à la* Geller, summon up spirits *à la* Doris Stokes or just read the truth about some past cases of haunting, *Forbidden Knowledge* is in a class of its own. Lunar mermaids will hate it — but that's their problem. □

Steve Blinkhorn is Managing Director of PRD Ltd, 36-38 London Road, St Albans AL1 1NG, UK.

So, you want to run the lab?

Warren K. Sinclair

Handbook of Research Laboratory Management. By Virginia P. White. ISI Press, Philadelphia/Adam Hilger, Bristol: 1988. Pp. 240. \$49.95, £43.

IN MOST corporate enterprises there is the tacit assumption that even the youngest new employee at the foot of the ladder can climb the rungs and become the corporate head. It is expected that most employees will dream about their ultimate success in climbing this ladder, even though it will rarely be fully realized, and much of the administration of such an organization is based on that assumption.

Nothing so comparatively simple is possible in the administration of science. Although they are far from oblivious to personal recognition, most scientists do not care to occupy administrative posts themselves and sometime lack respect for those who do. Many chafe at administrative controls and even delight in circumventing them. In a well-run laboratory, research scientists tend to take for granted that supplies will always be available, that if they want a new pH meter it will be obtainable, that they (and their assistants) will get paid, that travel expenses to meetings will be met, graphs and slides will be prepared, medical problems will be covered by insurance, vacations will be ample and retirement secure, to say nothing of adequate space in which to work, to park a car and so on.

Who sees to all these things? They are seen to by the large cadre of workers in a big research laboratory ('the administration') who want to share in the exciting enterprise that is the laboratory's goal but can contribute only secondarily to its success. Because the main contribution must be from the scientists, everything in the institution has to be directed towards maximizing the creativity of those who make up the research staff. The institution has no other reason for being.

That sensible if rather obvious theme flows through Virginia White's *Handbook of Research Laboratory Management*. Wisely, the author does not attempt to put good administration before good science. However, the *Handbook* treats virtually every phase of its subject — from directors to research and administrative staff, the organization of research, the important problems of communication in science,

publication, safety, buildings, equipment and finance. Administrative services and how to run them are detailed. Even how to prepare for and run a scientific meeting is described in depth. Those responsible for purchasing or personnel, meetings or equipment, or library services could all learn something here about their job. And while the book mainly addresses the organization of large research laboratories, rather than university departments, its principles apply across the board. The author's emphasis is on practice in biological rather than physical laboratories, but the latter are not neglected.

Although many scientists do not care to be bothered about the details of administration, they do expect their own leaders to be scientists and (somewhat magically)



successful administrators too. No one who has just stepped from the ranks of senior scientists to be the head of a large research division can react with other than dismay to find that not only has he accepted the leadership of the scientific enterprise of the institution (which he expected), but he has also become responsible for all the support people in the division, for the safety of all, for the building and for all the property in the division. He does, of course, have help in discharging the latter responsibilities, and the people concerned make up much of 'the administration' in a laboratory.

The advantages of scientists as laboratory and departmental or division directors are discussed in the book and, indeed, scientific leadership of a scientific enterprise is usually to be preferred. Regrettably, however, little distinction is made between laboratory and division directors. In big laboratories, the two roles are very different — the division head has much more control over scientific programmes and funding in his area than does the laboratory director, whose field is broad policy and the overall welfare of the entire organization. Alexander Hollaender was the quintessence of a success-

ful division director of biology — at Oak Ridge National Laboratory (ORNL) — in the halcyon days of the Atomic Energy Commission (AEC). The author cites him frequently throughout, but oddly she does not refer to the vital role of the division director in relating to programme personnel at the headquarters (of the AEC) in Washington. This key function involves representing the programme and the new work of the laboratory division to the funding agency in both informal and formal ways. Hollaender was superb in this role and his research division flourished accordingly.

Three aspects of research staff management that Virginia White deals with deserve further comment. The first is her emphasis on the importance of involving

researchers in every administrative proposal. Consultation with the research staff is vital to a successful laboratory. Good administrators know this; others can learn it here. The next concerns fallow periods in research workers' careers. Staff researchers in big laboratories, surrounded by able technical support, respected for past achievements and prominent in their favourite scientific society, nonetheless sometimes drift and need a new stimulus. Sabbaticals can be important but there are other ways, such as contact with students, especially graduate students, and teaching assignments at nearby universities.

The final aspect is the research support staff, who are invaluable in any laboratory. They are often longer-term employees than the research staff, develop skills in a lifetime of technical work that are critical to successful experiments, fix faltering equipment and generally to do all the 'man-Friday' work. Big laboratories tend to be excellent employers of such research assistants — they provide suitable titles, good working environments and better pay than most university programmes, where the problems experienced by non-faculty support staff are legend.

Omissions from the book are few, although some reference to bothersome items, such as security and quality assurance, and a more comprehensive treatment of radiation safety, would have been useful. The author ends with comments on the declining number of students entering science in the United States. This topic deserves much more discussion than is possible here but mention of it serves to point the finger at a serious problem. Equally serious may be the loss, due to budget cuts, of important areas of scientific expertise, for example in experimental carcinogenesis, in once broadly based

laboratories of the type Hollaender used to direct.

The author draws heavily on her own experience at ORNL, at the Salk Institute and elsewhere but she also quotes widely from the experiences of many others in laboratory administration; indeed, the refreshing quotations to be found throughout the text relieve occasional tedium. These days, the administration of scientific research is a very special field. Virginia White brings out the elements of

this speciality and provides a manual on how to do it well. Her *Handbook* will be a welcome guide to many, a substantial resource for others and interesting reading for many past and present laboratory administrators and research staff. □

Warren K. Sinclair is President of the National Council on Radiation Protection and Measurements, 7910 Woodmont Avenue, Suite 1016, Bethesda, Maryland 20814, USA. He was previously Associate Laboratory Director of Argonne National Laboratory, Argonne, Illinois.

Minding business

Stuart Sutherland

Molecules and Minds: Essays on Biology and the Social Order. By Steven Rose. Open University Press: 1988. Pp. 144. Hbk £25; pbk £8.95.

STEVEN ROSE lives in a world of bogeymen and Aunt Sallies. He howls at reductionism, sociobiology, the bourgeoisie and determinism. In so far as his collection of essays has any general thesis, it is that current scientific work supports the capitalist *status quo*, which ought not to be supported; rather it should give way to some undefined form of socialism that would produce — all evidence to the contrary — Utopia.

According to Rose, "reductionism . . . claims that in the long run higher order levels will be collapsed into the lower order ones; that they are 'nothing but' particularly complex systems to solving the problem of which the equations of the physicist are slowly approaching" (like many of his sentences, this one may need reading twice). Here as elsewhere he attacks something that no one, or almost no one, believes. The truth is that we will always require concepts appropriate to higher levels, no matter how much they are reducible. The hardness of a table top arises as an emergent quality from the way the atoms composing it are arranged and doubtless Steven Rose's angry writing arises from the way the neurons in his brain are arranged. But even if we fully understood his brain we would still need the term 'angry' to describe his behaviour. The reason why we cannot understand higher levels wholly in terms of lower ones lies in the nature of our own minds. We can consciously manipulate only a few concepts at once and, even though the table top is nothing but a vast collection of atoms, we cannot conceptualize its properties merely in terms of atomic activity; hence we need such concepts as hardness.

Having got reductionism all wrong, Rose suggests that it supports capitalism. The only instance he gives is that reductionism leads to putting the (lower-order) individual before the (higher-order)

society, which he thinks is a bad thing though many would disagree with him.

Another of Rose's favourite targets is sociobiology. He condemns the whole subject, despite its brilliant account of the behaviour of ants and other social insects, but here his main quarry is Edward Wilson. When scientists, including Rose, stray from their field, they often say foolish things and Wilson is no exception, for, having gone to the ant and studied her ways, he has not become wise. Rose attacks him for his tongue-in-cheek belief that a society in which sex differences were deliberately enhanced through conditioning would be happier than the present one. Without trying the experiment neither Rose nor anyone else has any means of knowing what the outcome would be. Oddly, he fails to pinpoint the two basic flaws in Wilson's popular writings. First, applying sociobiology to human affairs is for the most part pure speculation because little is known for certain about human genetics. Second, no amount of knowledge of that subject would tell us what is moral, though it might determine what is feasible.

It is unclear what Rose has against the much maligned bourgeoisie. Apparently they dominate others in rather a nasty way, but few will have noticed them doing this, unless of course voluntary work is a subtle form of domination. They make a large contribution to science, the arts, medicine and so on, but Rose is blind to their virtues. Nor does he stop to consider whether in his socialist Utopia a new bourgeoisie might arise, though if he were in charge he would doubtless take extreme measures to prevent it, for his book is redolent of Big Brother. For example, he would tell all scientists what to work on, though the projects would — he hopes — be of social benefit. He implies that he would also make people live healthy lives, though it seems doubtful if abstaining from meat and dairy products, not to mention running a regimented ten miles before breakfast, would appeal to all members of the new Utopia. But then what others want is irrelevant — all that matters is what Rose thinks they ought to want.

Because he does not define the term, Rose's views on determinism cannot be

explicated. Although he thinks it a bad thing, he seems to believe that behaviour is fully determined by the state of the brain. But then that is what he works on himself, as he describes in his final chapter which is almost the only one not given to obfuscation. His work seems to have affected his views on yet another moral issue. Because he gives both foot shock and electroconvulsive shock to animals, the Animal Liberation Movement is almost the only revolutionary campaign in Britain with which he does not side.

Rose not merely condemns the *obiter dicta* of scientists, many of which are admittedly rather silly, he actually believes that they influence politicians. He finds it hard to defend this thesis because the only examples he provides of politicians citing scientific evidence are the National Front and Enoch Powell. In fact, apart from the technological spin-off, work in biology and psychology has had remarkably little impact on political decisions. For example, at a time when most psychologists thought IQ was mainly inherited, the American government launched Headstart, a programme to improve the IQ of disadvantaged children. Again, feminism arose and received considerable sympathy at just the time when some sociobiologists were telling women their place was genetically determined and they should stick to it. The thought of Margaret Thatcher sitting up at night reading the works of E. O. Wilson, Robert Trivers and John Maynard Smith is not without appeal, but one suspects it is unrealistic.

Even Rose's debating points don't seem to work. He writes, "Genes for altruism says Richard Dawkins are any that confer upon their possessors the property of acting 'altruistically'". Rose comments "Genes for bad teeth in carnivores would result in their reducing their share of any common food resource to the benefit of others". He goes on to suggest it is absurd to define "bad tooth genes" as altruistic ones, having failed to notice that having a bad tooth is hardly an action.

Despite Rose's outcry, the truth is that science shapes neither morality nor government policy except in so far as new technologies can be applied. Nor, as Rose alleges, is science a particularly capitalist institution. Indeed capitalism itself is not necessarily a bad thing — it is hard to think of any other society in which the state, rightly or wrongly, has been more caring of its citizens than contemporary Britain. One suspects that people being what they are, there is no recipe for the perfect society for which Rose hankers. He might have done better to describe it rather than engage in the easier task of knocking down a few Aunt Sallies and attacking some particularly ineffectual bogeymen.

Stuart Sutherland is Director of the Centre for Research on Perception and Cognition, University of Sussex, Brighton BN1 9QG, UK.

Southern comforts

George A. Llano

International Research in the Antarctic. Edited by Richard Fifield. Published for SCAR and the ICSU Press by Oxford University Press: 1988. Pp. 146. £25, \$45.

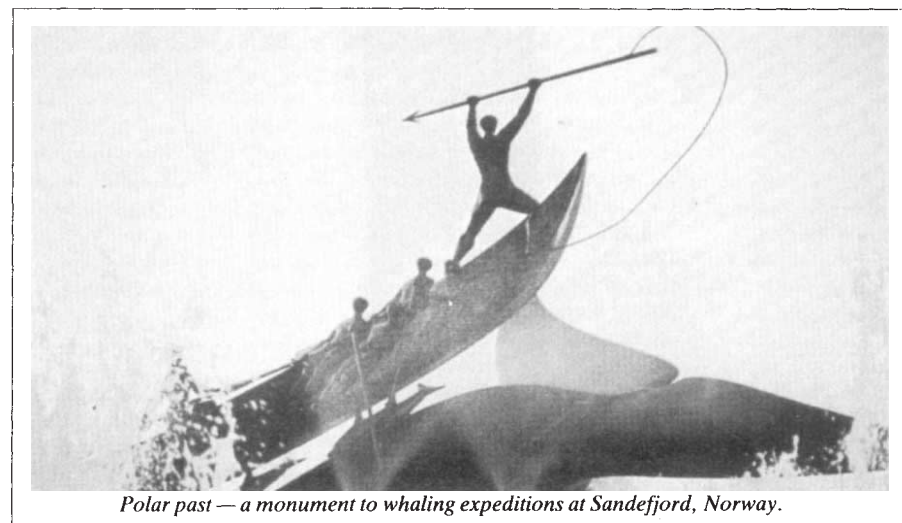
Antarctic Science. Edited by D.W.H. Walton. Cambridge University Press: 1987. Pp. 280. £25, \$39.50.

THIRTY years ago, those swept up by the excitement of plans for the 1957–1958 International Geophysical Year (IGY) sought and found little hard information on Antarctica. Despite the heroic record of exploration and hard-gained observations, the White Continent remained a blank space at the bottom of the world.

Following IGY, news correspondents provided the public with reports that at best gave an imperfect view of the scientific work that had been carried out, its objectives and how it had been planned. But for all the general lack of awareness of the fruits of IGY, cooperative endeavours were consolidated and extended under the multi-flag aegis of the Antarctic Treaty, a child of IGY, which was drawn up in 1959 and came into force in 1961. From 1957 onwards, research in the Antarctic has constituted an international success story. The continent is now the subject of a huge literature and a generation of scientists has been trained there. As Earth's scientific outpost, free from political or other outside interference, it has become identified as the Continent for Science.

The two books reviewed here cover scientific activity in Antarctica from IGY to the present. *International Research in the Antarctic* describes the structure and function of the eight permanent working groups of the Scientific Committee on Antarctic Research (SCAR), the coordinating body for research work. Details in the text about Antarctica itself, together with references to the instruments and methods used by scientists, are further clarified by appropriate illustrations. The final chapters — dealing with medicine, the environmental effects of man and freedom of exchange of information — are followed by two helpful appendices on SCAR publications and the Treaty, which conclude the book. Credit is due to the editor (or 'conductor', as he calls himself), Richard Fifield, for putting together this concise, clearly written and timely record. It meets a need that has existed since IGY for a popular but authoritative reference work on the means by which international scientific programmes evolve under the Antarctic Treaty.

The hallmark of the Treaty is *collaboration*, with free exchange of data and pooling of resources for large projects. Such has been the success of the system, which



Polar past — a monument to whaling expeditions at Sandefjord, Norway.

has generated huge amounts of new information, it has been hard to appreciate overall what has been done. All that information has now been brought together in *Antarctic Science*, an informative and readable account of what modern science has achieved on and above the continent.

The book starts with an introduction by Vivian Fuchs, a working scientist who played an important role in the development of the British Antarctic Survey (BAS), which is followed by five impressive sections each written by a BAS member. Part I sets the background by describing the events that led to the coming of age of science in Antarctica. Mention of whaling and the *Discovery* expeditions recalls the substantial bank of data on southern marine life, inherited from the British research on whales that began after the First World War.

Parts II, III and IV are the core of the book. Here, biological and physical phenomena are discussed in detail and there are descriptions of the pertinent research activities and findings. In "Life in a Cold Environment", Part II, marine and terrestrial organisms are contrasted, and their various adaptations examined; productivity, the intricacies of terrestrial, freshwater and marine food webs, and the need for the conservation of living resources are the main themes. Climate has a prominent place in both Parts III and IV. The former deals with the ice sheet, the underlying bedrock and associated matters, while the latter covers the Antarctic atmosphere and geospace. During IGY, these general research areas were well funded; international cooperation made the use of sophisticated instruments possible, and in part because Antarctica affords a unique platform for such work, a truly extraordinary range of studies on and over the Earth was carried out. The most recent finding was the discovery of the austral-spring ozone hole over Antarctica which, other implications apart, told the world that science in Antarctica is working.

Part V of *Antarctic Science* looks to the future. The research effort has always been unequally divided between the Treaty nations, but recently there has arisen the suspicion that sharing of natural resources rather than scientific research has become the main motive of the 'Club', the original Treaty signatories. Non-member Third World countries unable to meet the Treaty requirements have called for a takeover by the United Nations to make Antarctica part of the "global commons", while some 150 organizations from 37 countries, represented by the Antarctic and Southern Coalition, and including the International Union for the Conservation of Nature and Natural Resources, the Humane Society and the World Wildlife Fund, would set back the clock by designating Antarctica a protected wilderness, in effect giving the continent back to the Ice Ages.

These two books complement each other. They are a powerful argument in favour of the Treaty system which, though far from perfect, is preferable to anything else on offer. *International Research in the Antarctic* tells how the Antarctic experiment worked with SCAR in an advisory role to the Treaty of 1959, and how science developed in no man's land. *Antarctic Science*, richly illustrated with photographs and ingenious diagrams, documents the cornucopia of knowledge that has come from the joint efforts of scientists of many nationalities.

The message of the second book is to be found in the title of the sober yet searching final chapter, in which Richard Laws discusses current events and what they portend for science in Antarctica. "Cooperation or Confrontation?" — we could well get the answer in 1991, when, if a consultative party so requests, the Antarctic Treaty may be reviewed. □

George A. Llano, 450 Palm Circle West, Naples, Florida 33940, USA, was Program Manager of the Polar Biological and Medical Sciences in the Division of Polar Programs, National Science Foundation, 1961 to 1977.